PLANE CRASHES SINCE 1908

FINAL PROJECT

12/19/2022

Team D - Andrew Alford, Ryan Summers, Bill Steel, Anand Louis

PLANE CRASHES SINCE 1908

Table of Contents

Introduction

Data Cleansing/Munging

Analyses

Conclusions

Appendix

- R-code
- Raw Data Sets

INTRODUCTION

In today's modern world, airplanes are the most common mode of transportation. A single airplane crash can lead to significant losses. This report presents the findings of the airplane crashes and fatalities dataset with approximately 5,000 incidents from 1908 to 2019. The purpose of this report is to uncover and extract patterns to provide actionable insights. To gain a deeper understanding of the dataset, our team performed data cleansing, munging, and visualization techniques. Additionally, we performed text mining to identify the most common reasons for crashes from the summary section of the dataset. As a result of our analyses, our team will provide actionable insights into the crashes so they can be studied further to prevent future accidents.

RESEARCH QUESTIONS/ANALYSES:

1. How have plane crashes varied over time?
2. Which airline operator has the most crashes?
3. How have crashes varied between civilian and military over time?
4. What are the most common reasons for the crashes?
5. What US states have seen the most crashes over time?
6. How have fatalities have varied over time? Is safety improving?

PLANE CRASHES SINCE 1908

Table of Contents

Introduction

Data Cleansing/Munging

Analyses

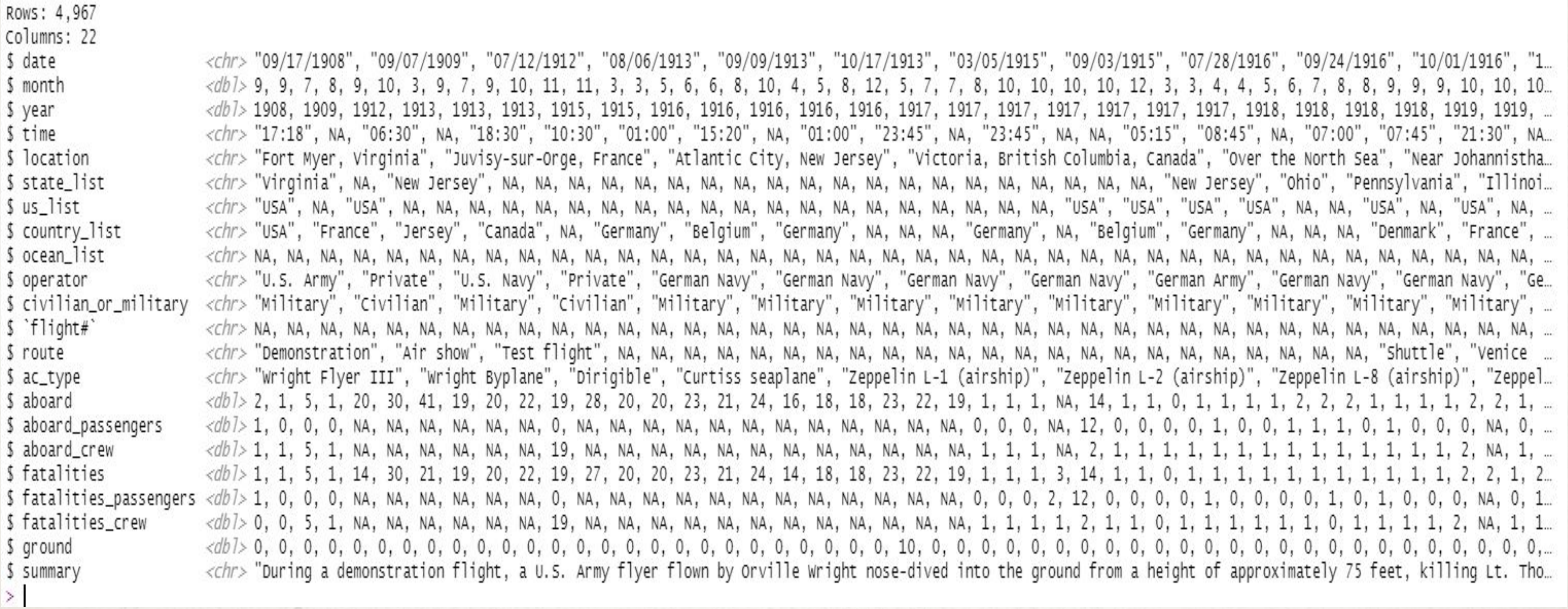
Conclusions

Appendix

- R-code
- Raw Data Sets

DATA CLEANSING/MUNGING

Our raw data set data (From www.kaggle.com/datasets/cgurkan/airplane-crash-data-since-1908) has almost 5000 observations each one reflecting a plane crash since 1908. All but one of the 22 variables are characters with Fatalities being the only numerical variable. Missing data is very prevalent in the form of *NA*s or the word "NULL".



```
Rows: 4,967
Columns: 22
$ date                  <chr> "09/17/1908", "09/07/1909", "07/12/1912", "08/06/1913", "09/09/1913", "10/17/1913", "03/05/1915", "09/03/1915", "07/28/1916", "09/24/1916", "10/01/1916", "1…
$ month                 <dbl> 9, 9, 7, 8, 9, 10, 3, 9, 7, 9, 10, 11, 11, 3, 3, 5, 6, 6, 8, 10, 4, 5, 8, 12, 5, 7, 7, 8, 10, 10, 10, 10, 12, 3, 3, 4, 4, 5, 6, 7, 8, 8, 9, 9, 9, 10, 10, 10…
$ year                  <dbl> 1908, 1909, 1912, 1913, 1913, 1913, 1915, 1915, 1916, 1916, 1916, 1916, 1916, 1917, 1917, 1917, 1917, 1917, 1917, 1917, 1918, 1918, 1918, 1918, 1919, 1919, …
$ time                  <chr> "17:18", NA, "06:30", NA, "18:30", "10:30", "01:00", "15:20", NA, "01:00", "23:45", NA, "23:45", NA, NA, "05:15", "08:45", NA, "07:00", "07:45", "21:30", NA…
$ location              <chr> "Fort Myer, Virginia", "Juvisy-sur-Orge, France", "Atlantic City, New Jersey", "Victoria, British Columbia, Canada", "Over the North Sea", "Near Johannistha…
$ state_list            <chr> "Virginia", NA, "New Jersey", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "New Jersey", "Ohio", "Pennsylvania", "Illinoi…
$ us_list               <chr> "USA", NA, "USA", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "USA", "USA", "USA", "USA", NA, NA, "USA", NA, "USA", NA, …
$ country_list          <chr> "USA", "France", "Jersey", "Canada", NA, "Germany", "Belgium", "Germany", NA, NA, NA, "Germany", NA, "Belgium", "Germany", NA, NA, NA, "Denmark", "France", …
$ ocean_list            <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
$ operator              <chr> "U.S. Army", "Private", "U.S. Navy", "Private", "German Navy", "German Navy", "German Navy", "German Navy", "German Army", "German Navy", "German Navy", "Ge…
$ civilian_or_military  <chr> "Military", "Civilian", "Military", "Civilian", "Military", "Military", "Military", "Military", "Military", "Military", "Military", "Military", "Military", …
$ `flight#`             <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
$ route                 <chr> "Demonstration", "Air show", "Test flight", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, "Shuttle", "Venice  …
$ ac_type               <chr> "Wright Flyer III", "Wright Byplane", "Dirigible", "Curtiss seaplane", "Zeppelin L-1 (airship)", "Zeppelin L-2 (airship)", "Zeppelin L-8 (airship)", "Zeppel…
$ aboard                <dbl> 2, 1, 5, 1, 20, 30, 41, 19, 20, 22, 19, 28, 20, 20, 23, 21, 24, 16, 18, 18, 23, 22, 19, 1, 1, 1, NA, 14, 1, 1, 0, 1, 1, 1, 1, 2, 2, 2, 1, 1, 1, 1, 2, 2, 1, …
$ aboard_passengers     <dbl> 1, 0, 0, 0, NA, NA, NA, NA, NA, NA, 0, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 0, 0, 0, NA, 12, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, NA, 0, …
$ aboard_crew           <dbl> 1, 1, 5, 1, NA, NA, NA, NA, NA, NA, 19, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 1, 1, 1, NA, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, NA, 1, …
$ fatalities            <dbl> 1, 1, 5, 1, 14, 30, 21, 19, 20, 22, 19, 27, 20, 20, 23, 21, 24, 14, 18, 18, 23, 22, 19, 1, 1, 1, 3, 14, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 1, 2…
$ fatalities_passengers <dbl> 1, 0, 0, 0, NA, NA, NA, NA, NA, NA, 0, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 0, 0, 0, 2, 12, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, NA, 0, 1…
$ fatalities_crew       <dbl> 0, 0, 5, 1, NA, NA, NA, NA, NA, NA, 19, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 1, 1, 1, 1, 2, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 2, NA, 1, 1…
$ ground                <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 10, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,…
$ summary               <chr> "During a demonstration flight, a U.S. Army flyer flown by Orville Wright nose-dived into the ground from a height of approximately 75 feet, killing Lt. Tho…
> |
```

DATA CLEANSING/MUNGING

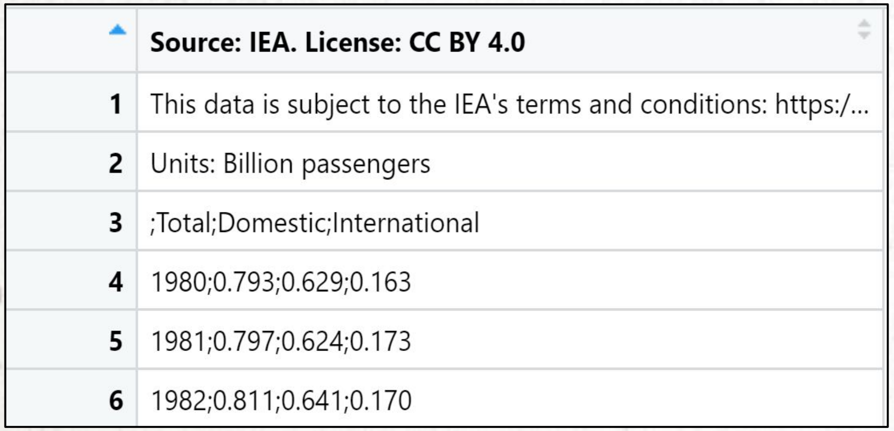
Cleansing and preparing the raw data for analyses was very involved and is summarized below:

1. Added 2 numerical columns from the "Date" Field. One is the Year, the other is the Month.
2. Took apart the "Location" field. Each word is separated by a "," which was used to create 4 more fields simply called loc1, loc2, loc3, loc4. This isolated anything within that field getting to state, country, city (when possible).
3. Changed to numeric all character fields that should be numeric.
4. Within the data set, several cells were brought in with the word "NULL". *NA* was forced into those cells to be consistent with what R does when a field is blank.
5. Removed the cn/ln column and the registration columns as they are unnecessary for our analyses.
6. Separated a text column with civilian vs military crashes into two firm "Civilian" vs. "Military" columns that are actionable.
7. Standardized column names in order to provide a firm set of data frames with which the team can do analyses.
8. Cleaned up the "Operator" column.
9. Created a new dataframe with crashes/fatalities by year.
10. Merged in a new raw data set (world passenger volume in billions since 1980).
11. Performed text extraction on the crash location field to get country/state/city/ocean.
12. Performed a text mining technique called N-grams, which extracts a sequence of text to break down the summary crash section of the data set. The process includes using stopwords, which eliminates unimportant words.

DATA CLEANSING/MUNGING

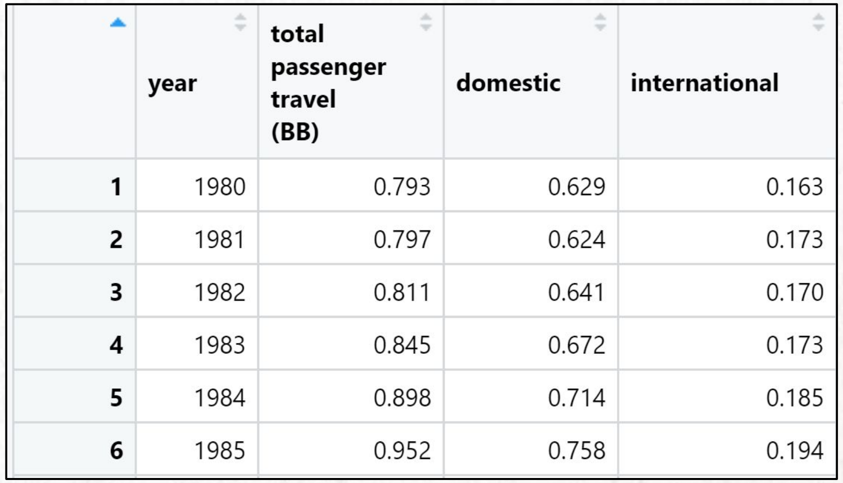
We also analyzed fatalities by year against total passenger volume. However, this data was not available in our original dataset. We found data available with passenger volume by year since 1980 (in billions domestically and internationally). This provided 40 years of data which is a good data set for various analyses. The raw data set needed cleansing. It originally came in as 1 column character string separated by semicolons. We parsed the data into several columns, added column names, and turned the values numeric.

Original Data



	Source: IEA. License: CC BY 4.0
1	This data is subject to the IEA's terms and conditions: https:/...
2	Units: Billion passengers
3	;Total;Domestic;International
4	1980;0.793;0.629;0.163
5	1981;0.797;0.624;0.173
6	1982;0.811;0.641;0.170

Cleansed Data

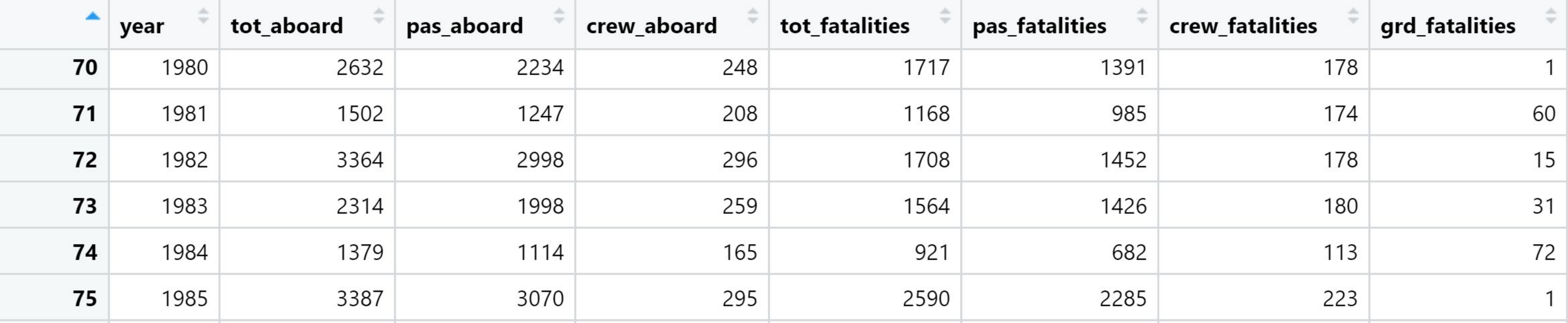


	year	total passenger travel (BB)	domestic	international
1	1980	0.793	0.629	0.163
2	1981	0.797	0.624	0.173
3	1982	0.811	0.641	0.170
4	1983	0.845	0.672	0.173
5	1984	0.898	0.714	0.185
6	1985	0.952	0.758	0.194

DATA CLEANSING/MUNGING

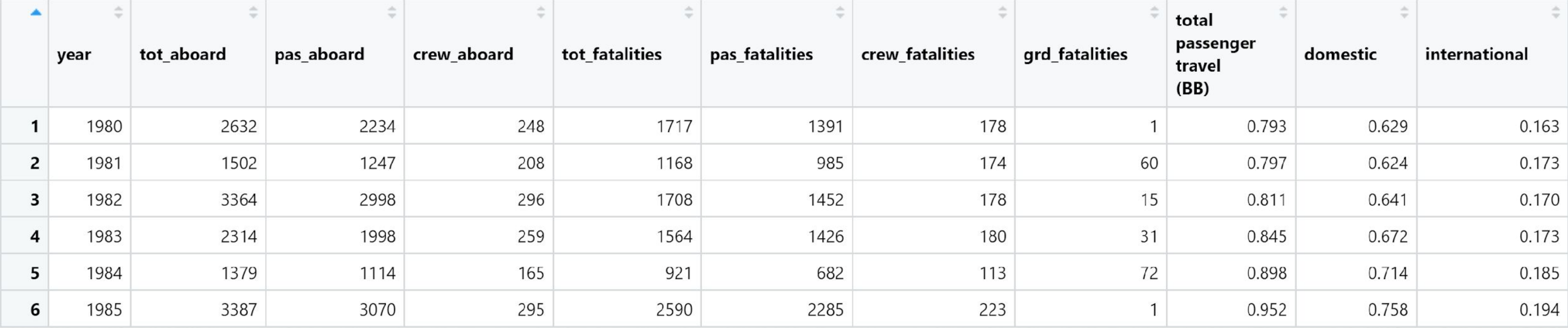
With passenger volume by year since 1980, and place crashes since 1908, we merged the data. Now we have a final data frame against which we can do analyses around crashes and fatalities/year given increases in passenger volume.

Summary:
Fatalities by year with relevant numeric columns



	year	tot_aboard	pas_aboard	crew_aboard	tot_fatalities	pas_fatalities	crew_fatalities	grd_fatalities
70	1980	2632	2234	248	1717	1391	178	1
71	1981	1502	1247	208	1168	985	174	60
72	1982	3364	2998	296	1708	1452	178	15
73	1983	2314	1998	259	1564	1426	180	31
74	1984	1379	1114	165	921	682	113	72
75	1985	3387	3070	295	2590	2285	223	1

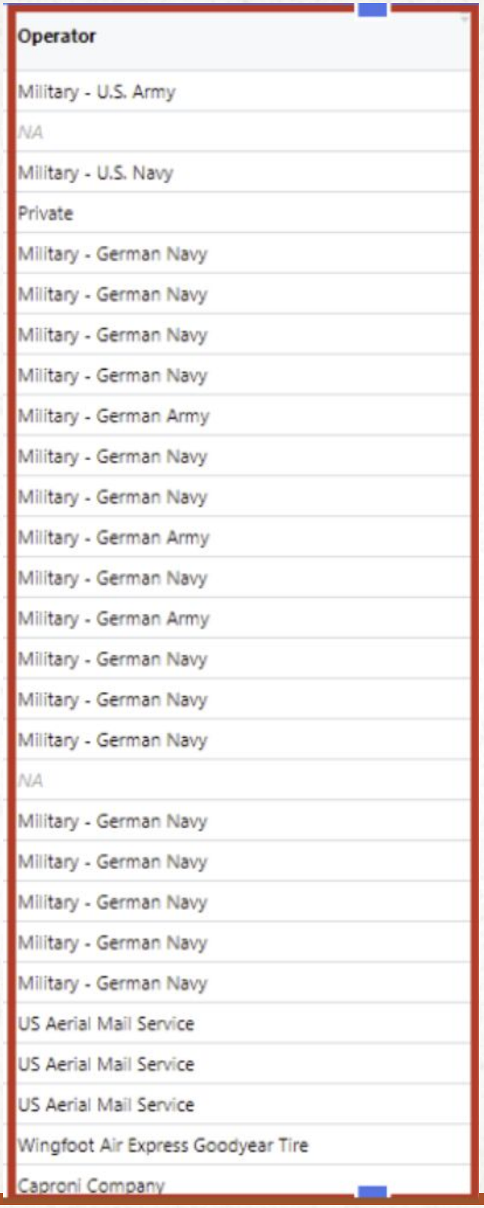
Merged data:
Fatalities by year with passenger volume



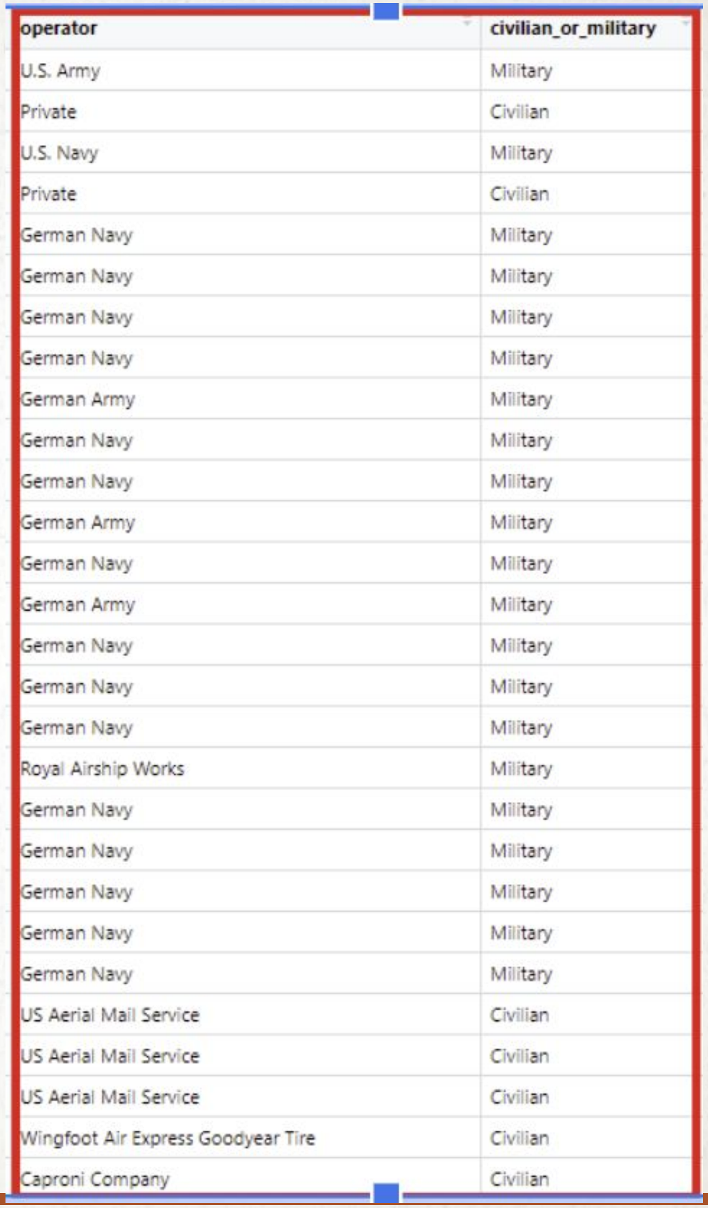
	year	tot_aboard	pas_aboard	crew_aboard	tot_fatalities	pas_fatalities	crew_fatalities	grd_fatalities	total passenger travel (BB)	domestic	international
1	1980	2632	2234	248	1717	1391	178	1	0.793	0.629	0.163
2	1981	1502	1247	208	1168	985	174	60	0.797	0.624	0.173
3	1982	3364	2998	296	1708	1452	178	15	0.811	0.641	0.170
4	1983	2314	1998	259	1564	1426	180	31	0.845	0.672	0.173
5	1984	1379	1114	165	921	682	113	72	0.898	0.714	0.185
6	1985	3387	3070	295	2590	2285	223	1	0.952	0.758	0.194

DATA CLEANSING/MUNGING

The process of munging the operator column was to first create a separate variable for the operator being either civilian or military so that the original operator column did not contain two data observations (i.e. both whether the operator was military or civilian and the name of the operator itself). Additionally, we filled in the NA values with data that could be extracted from the summary column.



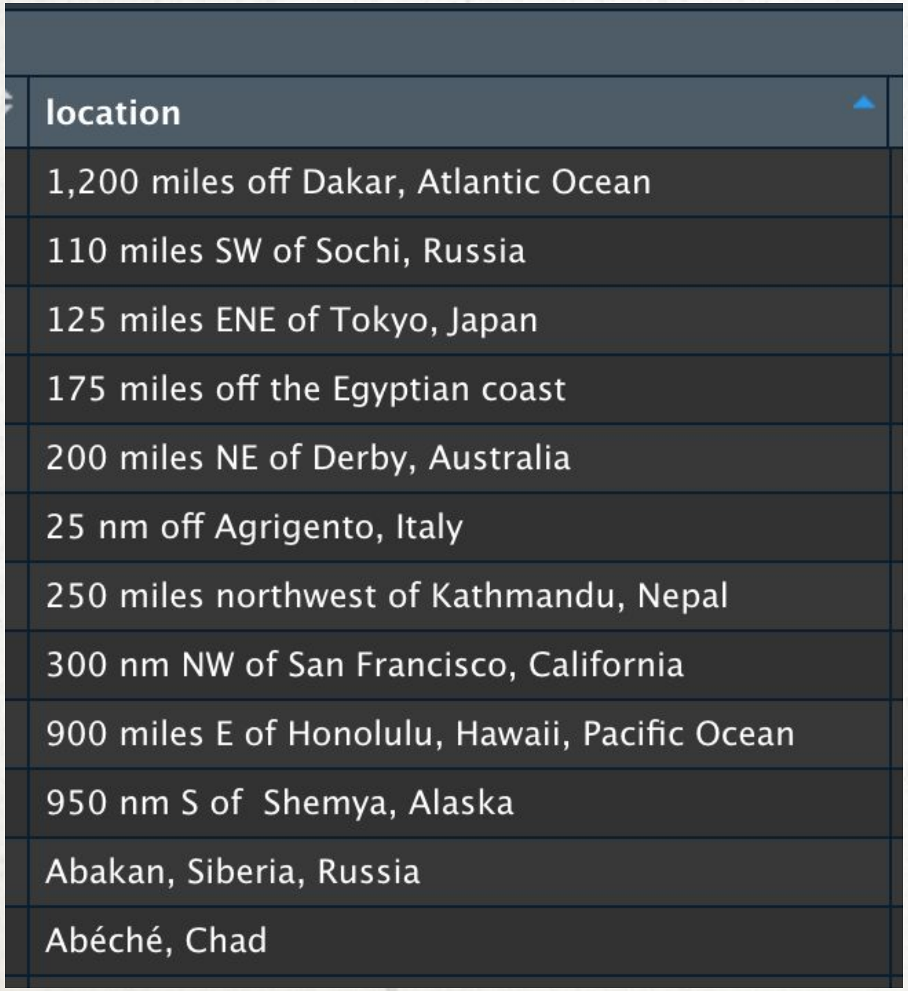
Operator
Military - U.S. Army
NA
Military - U.S. Navy
Private
Military - German Navy
Military - German Navy
Military - German Navy
Military - German Navy
Military - German Army
Military - German Navy
Military - German Navy
Military - German Army
Military - German Navy
Military - German Army
Military - German Navy
Military - German Navy
Military - German Navy
NA
Military - German Navy
Military - German Navy
Military - German Navy
Military - German Navy
Military - German Navy
US Aerial Mail Service
US Aerial Mail Service
US Aerial Mail Service
Wingfoot Air Express Goodyear Tire
Caproni Company



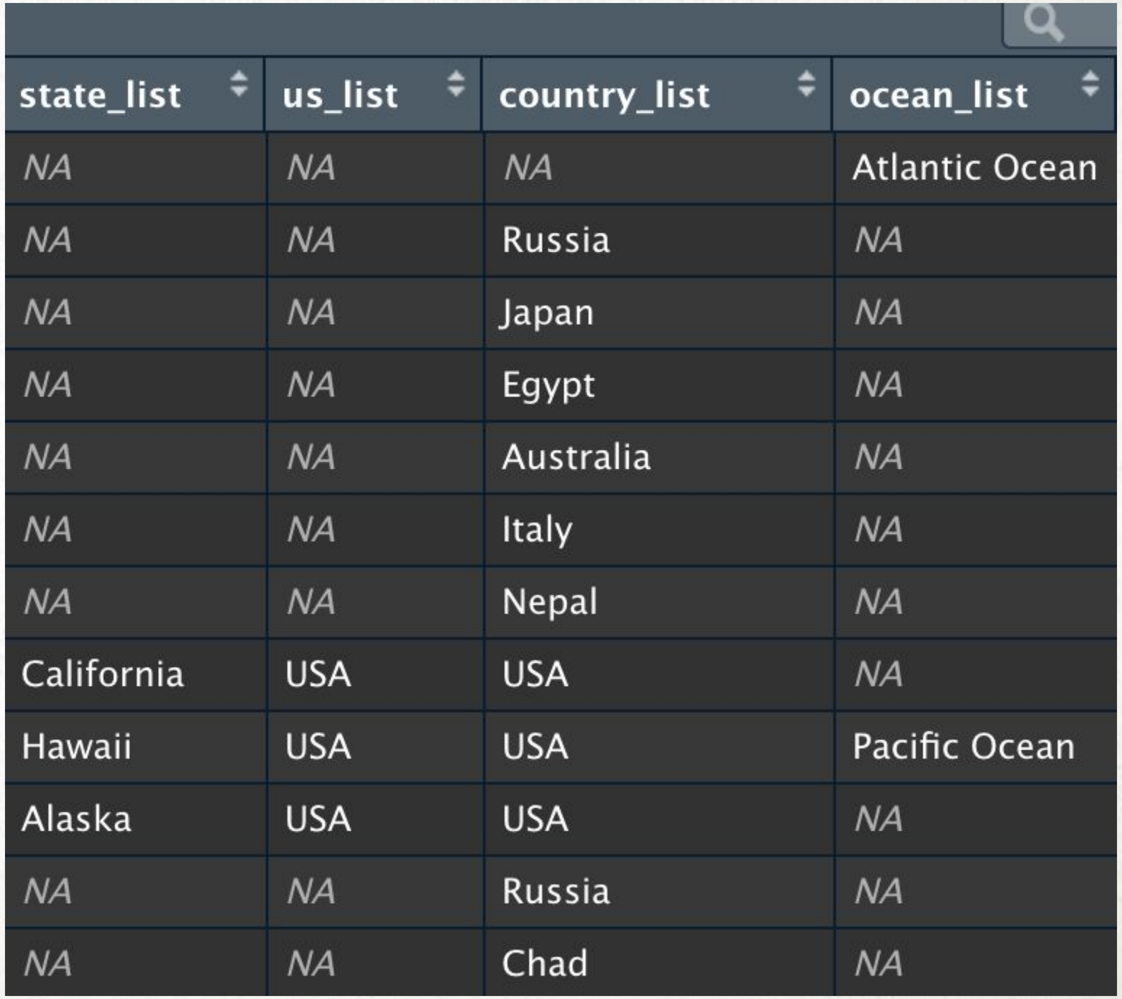
operator	civilian_or_military
U.S. Army	Military
Private	Civilian
U.S. Navy	Military
Private	Civilian
German Navy	Military
German Navy	Military
German Navy	Military
German Navy	Military
German Army	Military
German Navy	Military
German Navy	Military
German Army	Military
German Navy	Military
German Army	Military
German Navy	Military
German Navy	Military
German Navy	Military
Royal Airship Works	Military
German Navy	Military
German Navy	Military
German Navy	Military
German Navy	Military
German Navy	Military
US Aerial Mail Service	Civilian
US Aerial Mail Service	Civilian
US Aerial Mail Service	Civilian
Wingfoot Air Express Goodyear Tire	Civilian
Caproni Company	Civilian

DATA CLEANSING/MUNGING

We performed text extraction on crash location to distinguish between country, US state, and ocean.



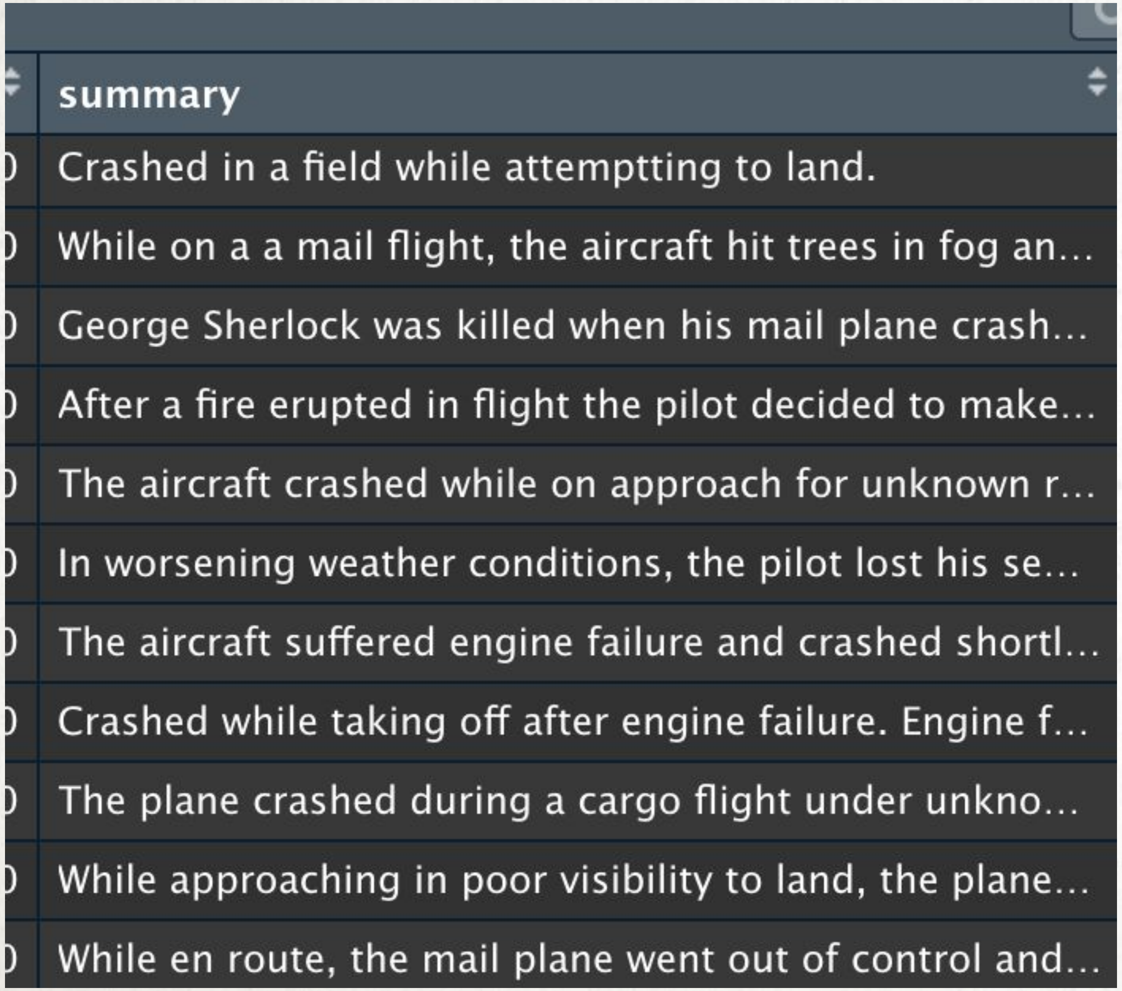
location
1,200 miles off Dakar, Atlantic Ocean
110 miles SW of Sochi, Russia
125 miles ENE of Tokyo, Japan
175 miles off the Egyptian coast
200 miles NE of Derby, Australia
25 nm off Agrigento, Italy
250 miles northwest of Kathmandu, Nepal
300 nm NW of San Francisco, California
900 miles E of Honolulu, Hawaii, Pacific Ocean
950 nm S of Shemya, Alaska
Abakan, Siberia, Russia
Abéché, Chad



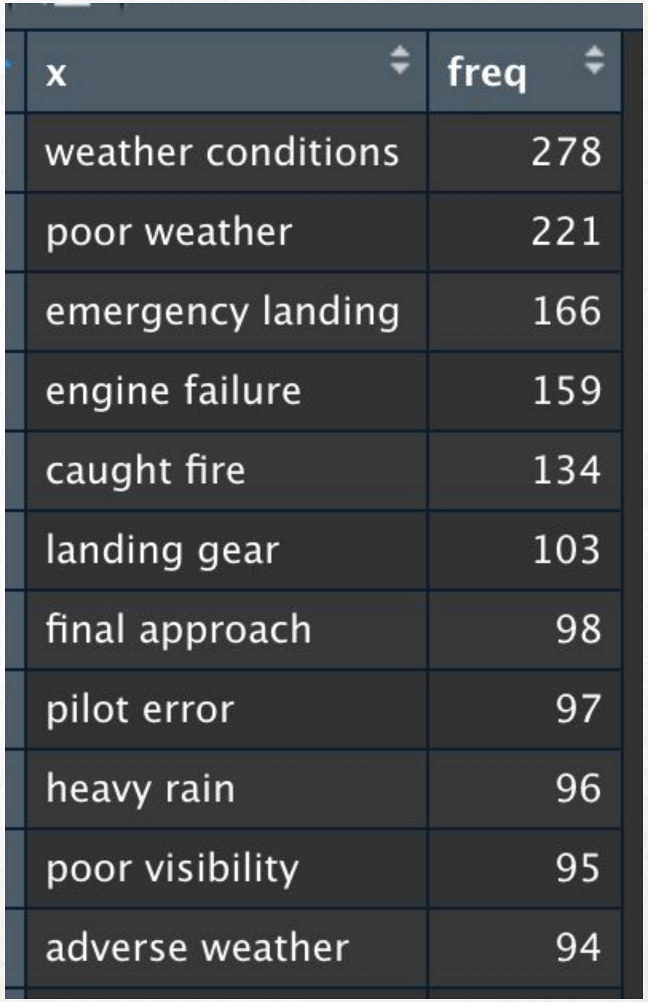
state_list	us_list	country_list	ocean_list
NA	NA	NA	Atlantic Ocean
NA	NA	Russia	NA
NA	NA	Japan	NA
NA	NA	Egypt	NA
NA	NA	Australia	NA
NA	NA	Italy	NA
NA	NA	Nepal	NA
California	USA	USA	NA
Hawaii	USA	USA	Pacific Ocean
Alaska	USA	USA	NA
NA	NA	Russia	NA
NA	NA	Chad	NA

DATA CLEANSING/MUNGING

We performed a text mining technique called N-grams, which extracts a sequence of text to break down the summary crash section of the data set. The process includes using stopwords, which eliminates unimportant words.



summary
Crashed in a field while attemptting to land.
While on a a mail flight, the aircraft hit trees in fog an…
George Sherlock was killed when his mail plane crash…
After a fire erupted in flight the pilot decided to make…
The aircraft crashed while on approach for unknown r…
In worsening weather conditions, the pilot lost his se…
The aircraft suffered engine failure and crashed shortl…
Crashed while taking off after engine failure. Engine f…
The plane crashed during a cargo flight under unkno…
While approaching in poor visibility to land, the plane…
While en route, the mail plane went out of control and…



x	freq
weather conditions	278
poor weather	221
emergency landing	166
engine failure	159
caught fire	134
landing gear	103
final approach	98
pilot error	97
heavy rain	96
poor visibility	95
adverse weather	94

PLANE CRASHES SINCE 1908

Table of Contents

Introduction

Data Cleansing/Munging

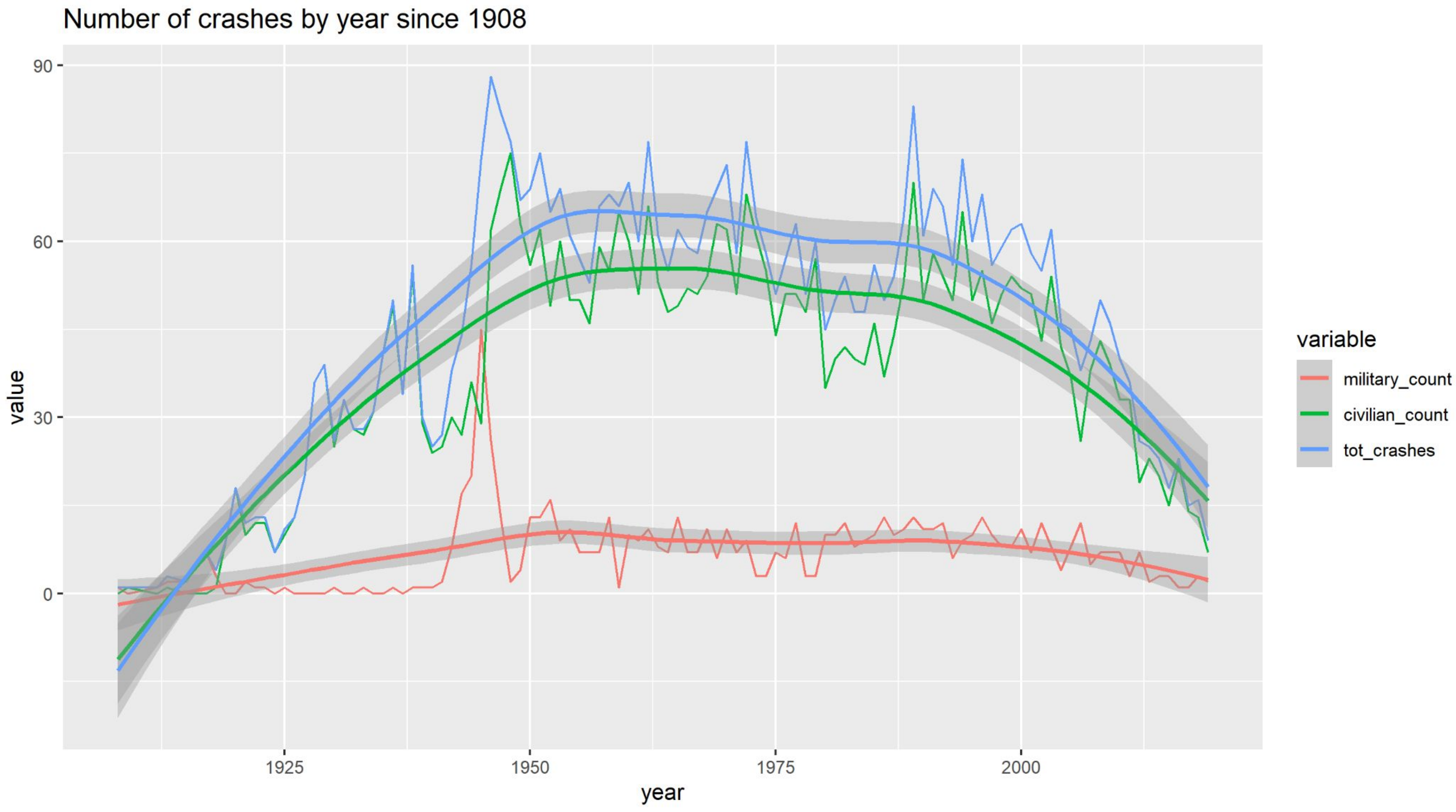
Analyses

Conclusions

Appendix
- R-code
- Raw Data Sets

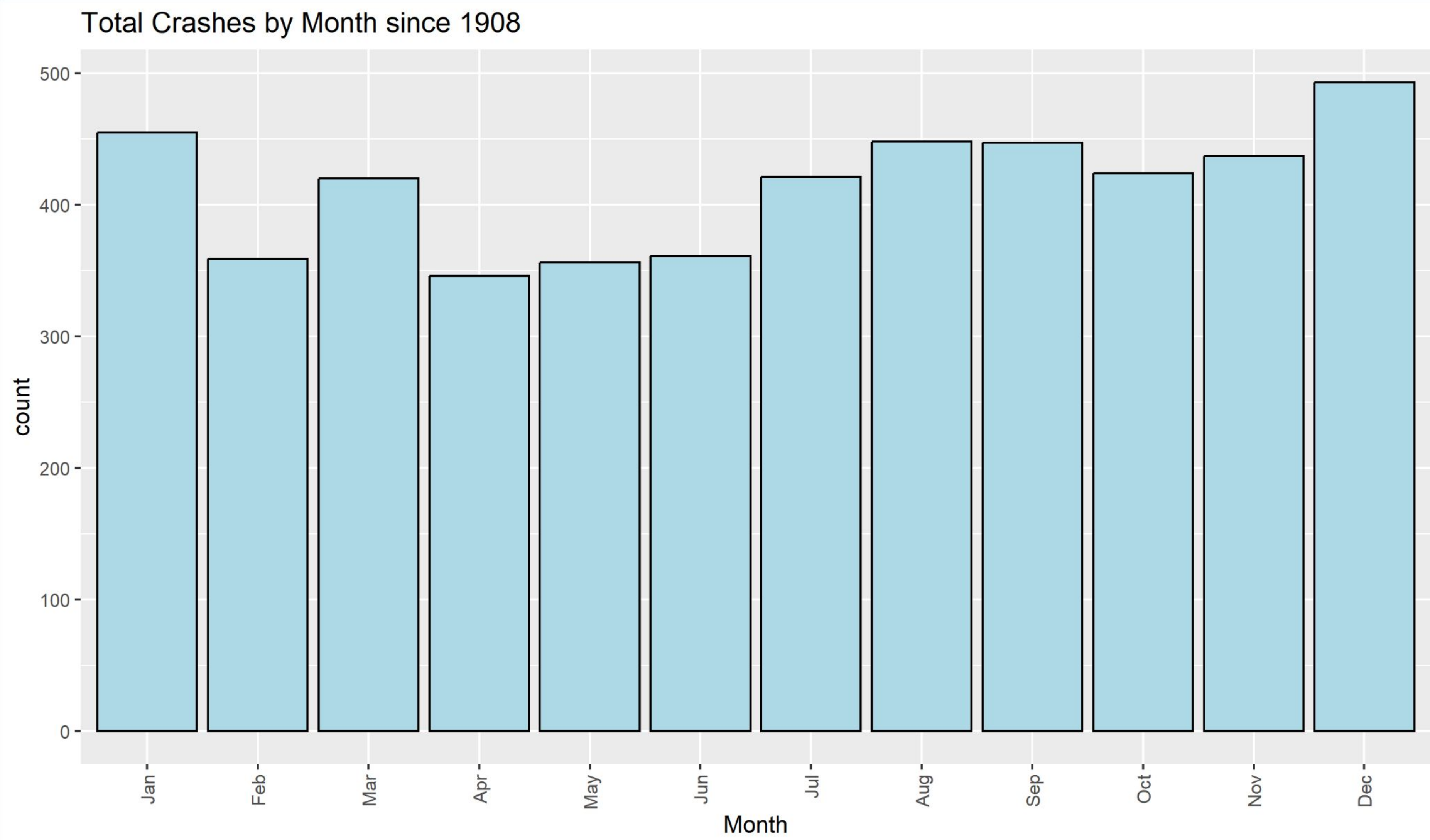
ANALYSES - General Crash Analysis

Crashes consistently increased each year from 1908 through the 1940's spiking during WWII which is understandable. From the 50's through the 70's, crashes held steady and started to decline from the 80's through present. The data below does not take increased passenger volume into account which will be looked at later.



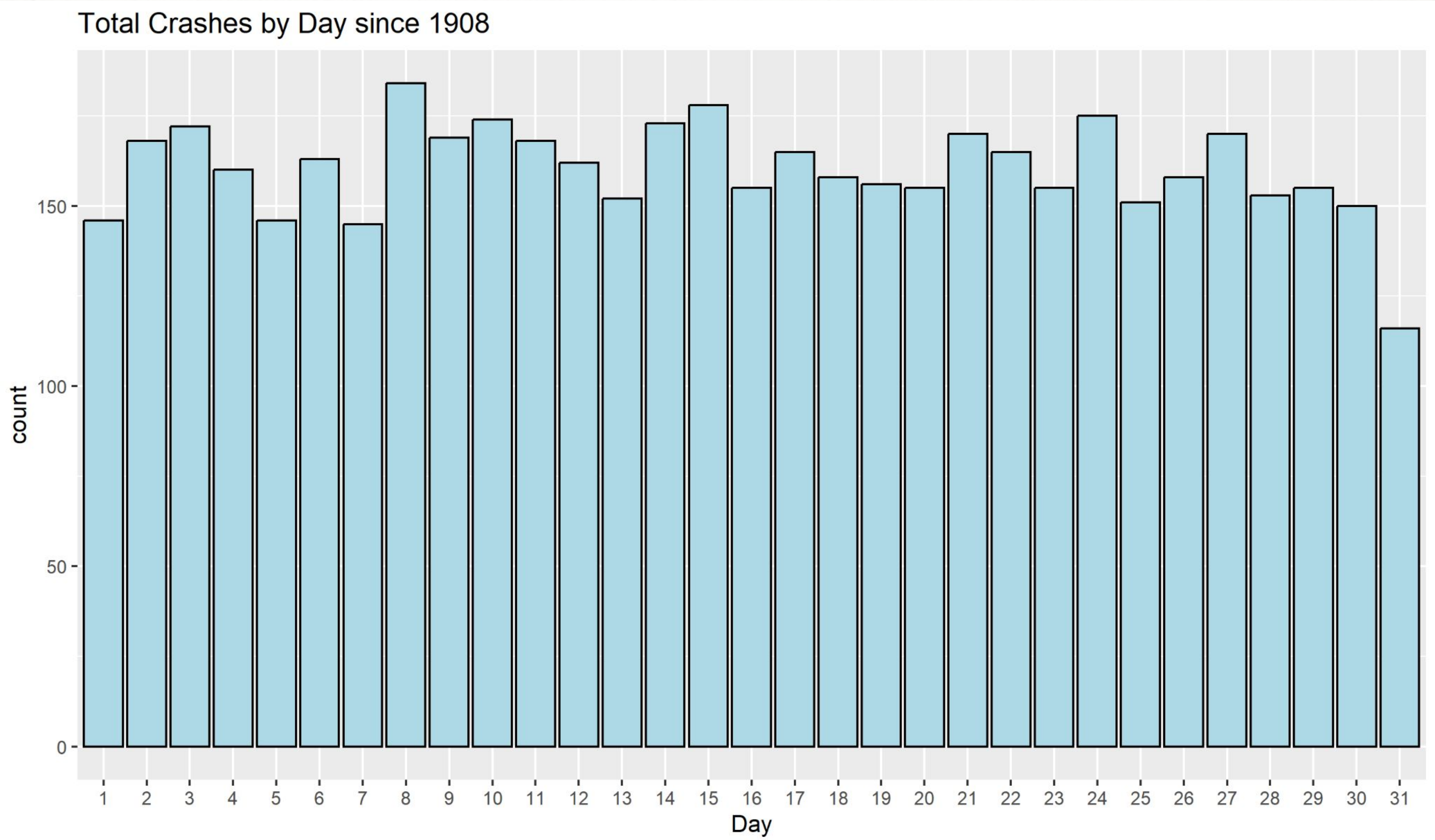
ANALYSES - General Crash Analysis

December was the month with the highest number of crashes followed by January. April/May were the months with the lowest number of crashes. Crashes were about 30% higher in December/January vs. April/May.



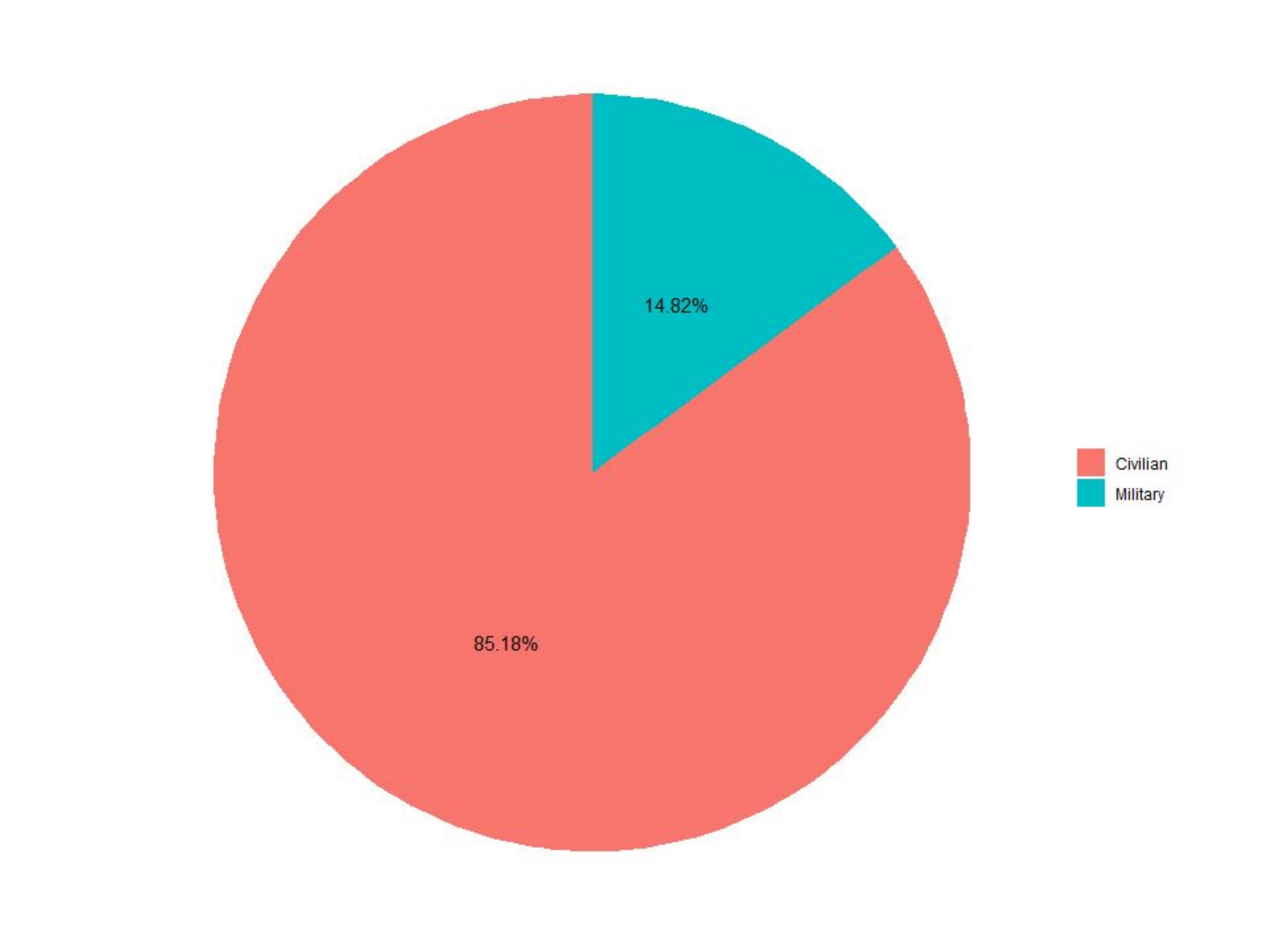
ANALYSES - General Crash Analysis

Crashes by day looks reasonably stable ranging between ~150-180 with the notable exception of the 31st which can be explained by 4 months not having a 31st day.



ANALYSES – Military vs. Civilian Crash Analysis

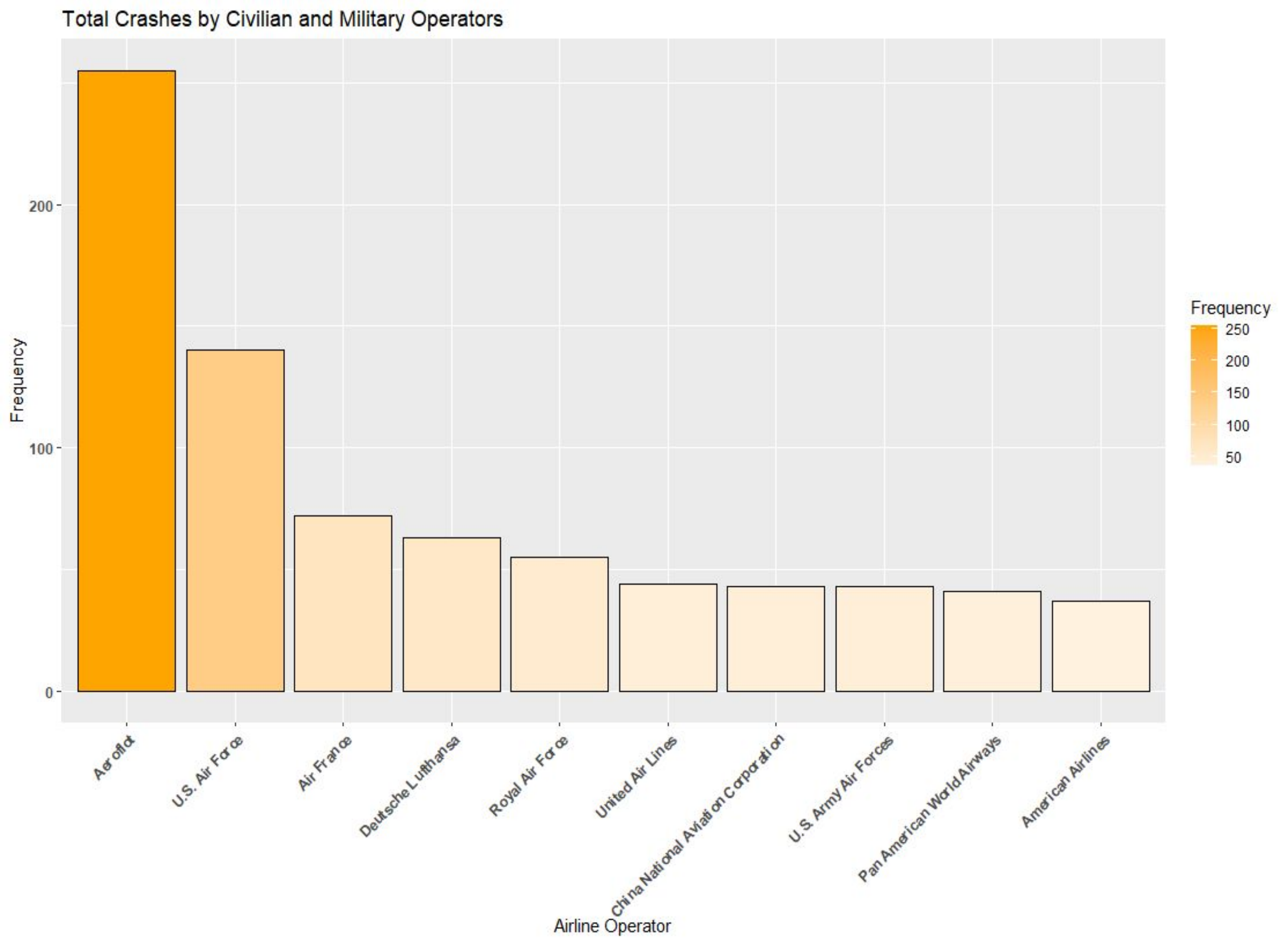
This pie chart shows the percentage breakdown of civilian vs military crashes. It is clearly evident from one's first impression that civilian crashes significantly outweigh the military crashes. The exact reason for this is largely unknown. It could perhaps be because there is a higher volume of civilian flights. Another reason could be military professionalism has higher standards and safety regulations.



ANALYSES – Military vs. Civilian Crash Analysis

This bar chart shows the top ten airline operators and air forces, both military and civilian, with the highest frequency of crashes in the data set. Clearly Aeroflot has far and away the highest frequency of crashes, followed by the U.S. Air Force. The cause of Aeroflot's high crash rate could potentially be explained by its long history as one of the oldest airlines, being founded in 1923. Aeroflot, a Russian airline, has had a notoriously bad reputation, at one time accounting for half of all airplane crash deaths worldwide.

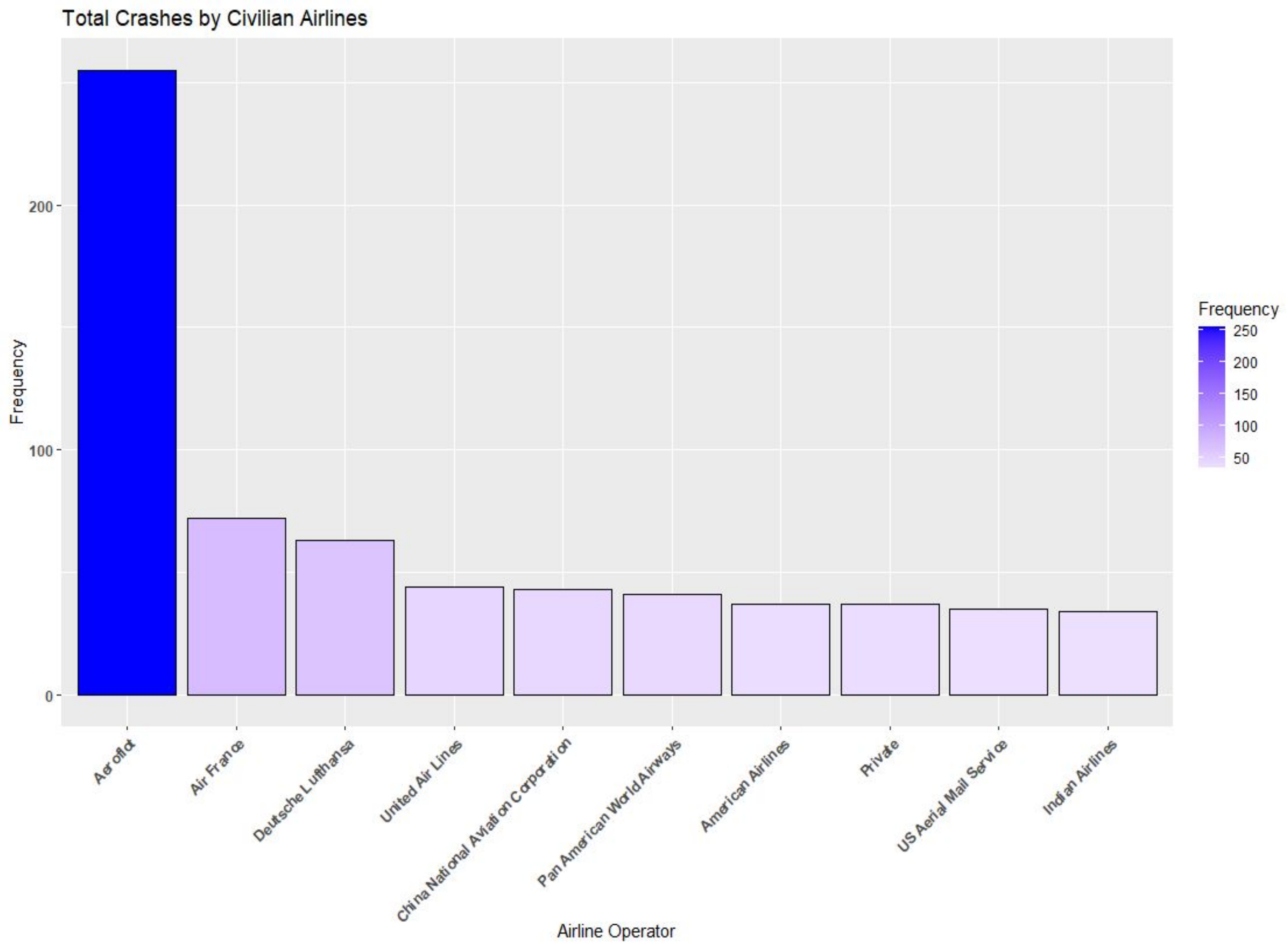
As one can see, these top two account for a significant amount of the total crashes, with the crash frequency generally stabilizing after the U.S. Air Force.



ANALYSES – Military vs. Civilian Crash Analysis

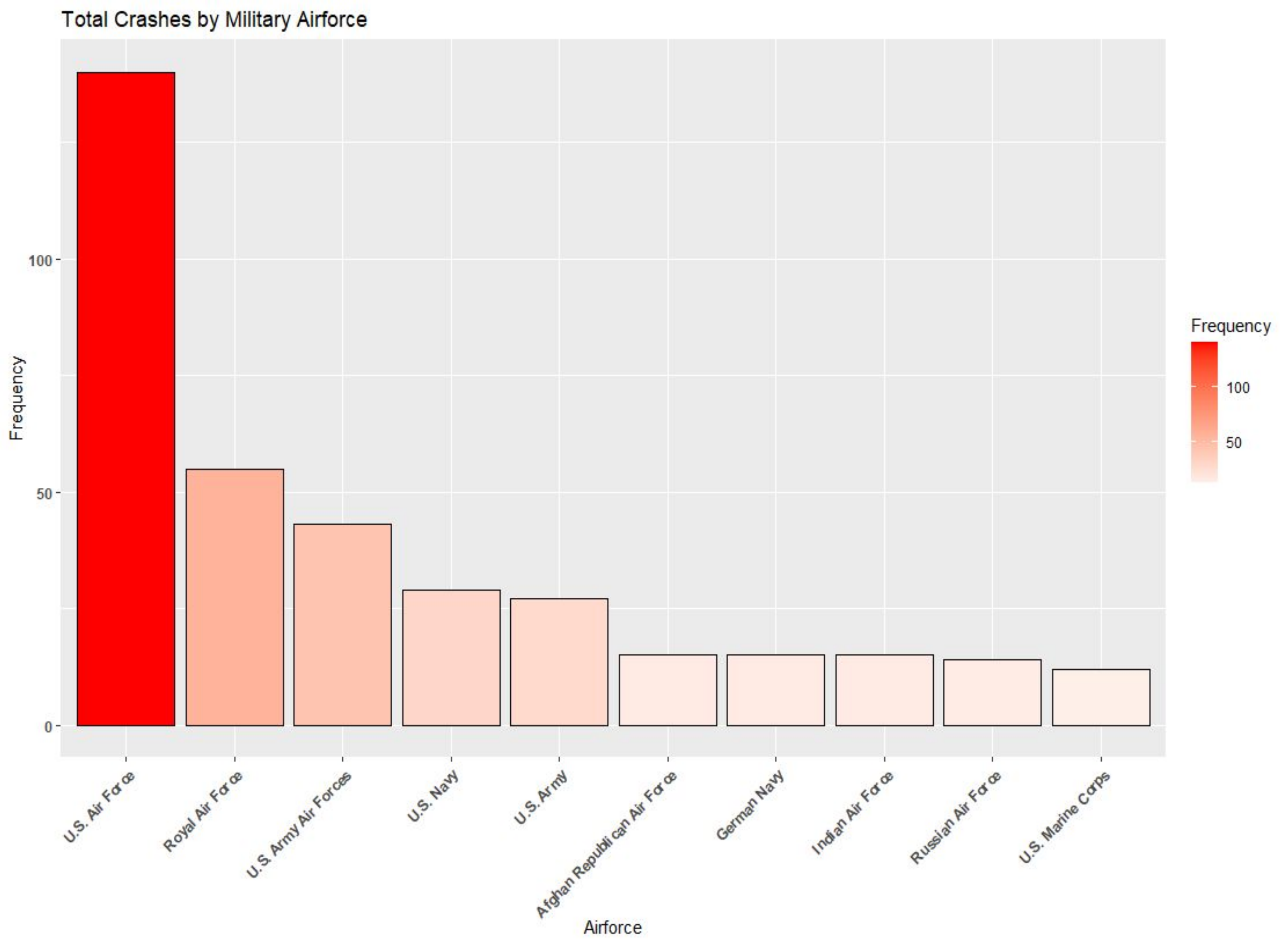
Similar to the bar chart previously presented. This chart displays the top ten airline operators, this time accounting for civilian airlines only. Aeroflot's status as the most dangerous airline is even more pronounced when compared to its civilian competitors.

One take away is to definitely be careful flying with Aeroflot!



ANALYSES – Military vs. Civilian Crash Analysis

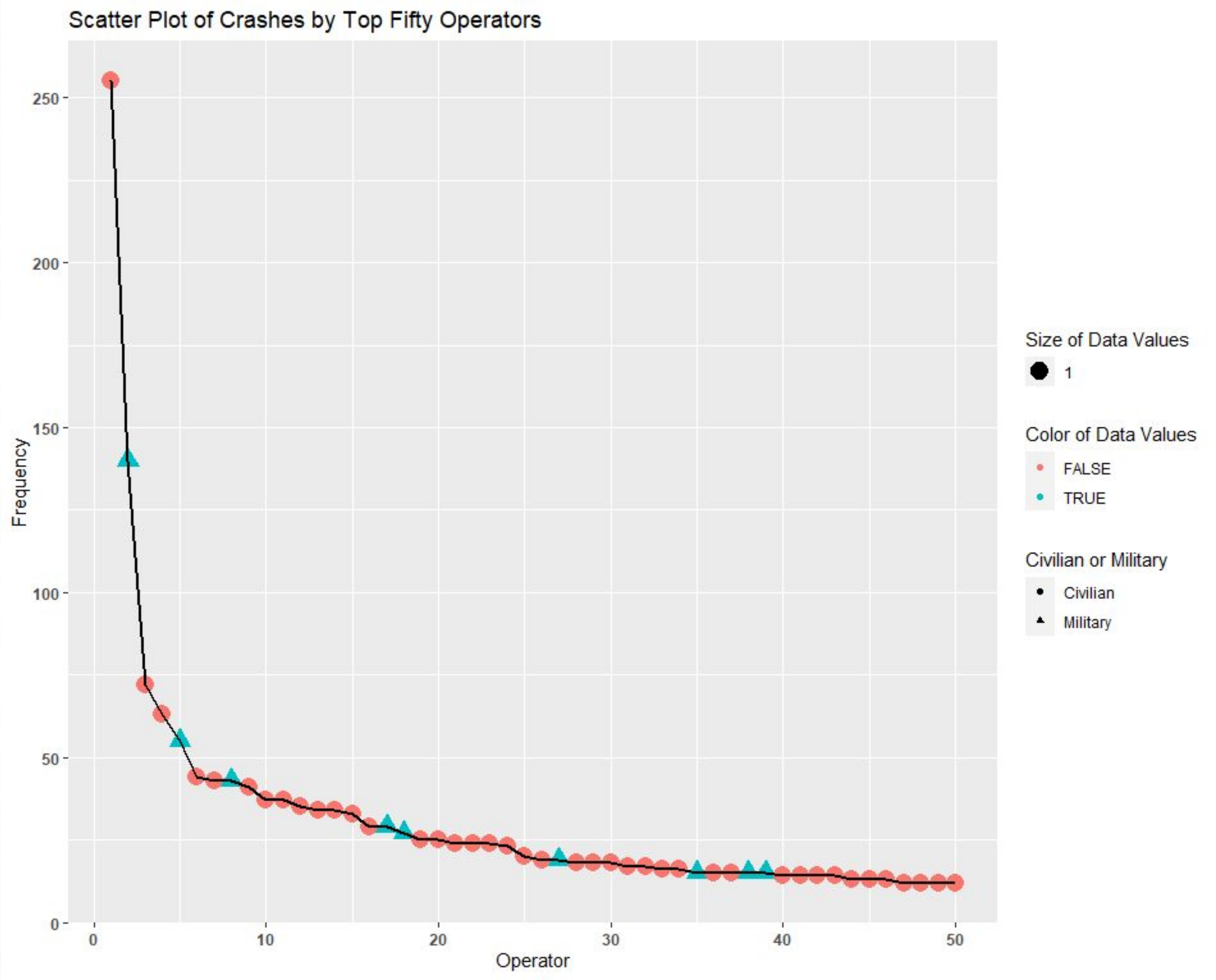
This bar chart shows the top ten air forces by crash frequency. One particular takeaway of this chart is the U.S. military's presence in the chart, accounting for half of the most dangerous air forces in the world, broken down by branch.



ANALYSES – Military vs. Civilian Crash Analysis

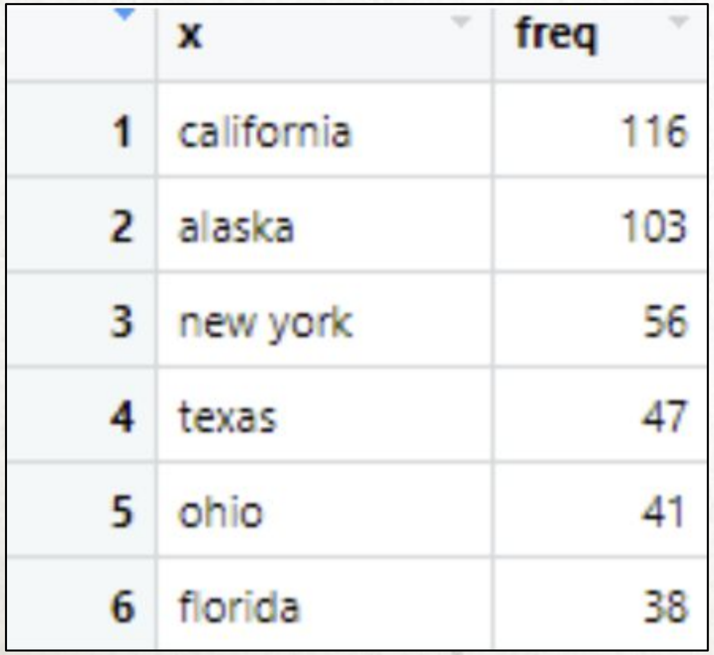
Though this chart is technically a Scatter plot, it is better to visualize it as a bar chart. The orange circles represent civilian airlines, and the green triangles represent military air forces.

The takeaway from this chart is that civilian airlines greatly outweigh military air forces in frequency of crashes, with 41 of the top 50 highest crashing operators being civilian.

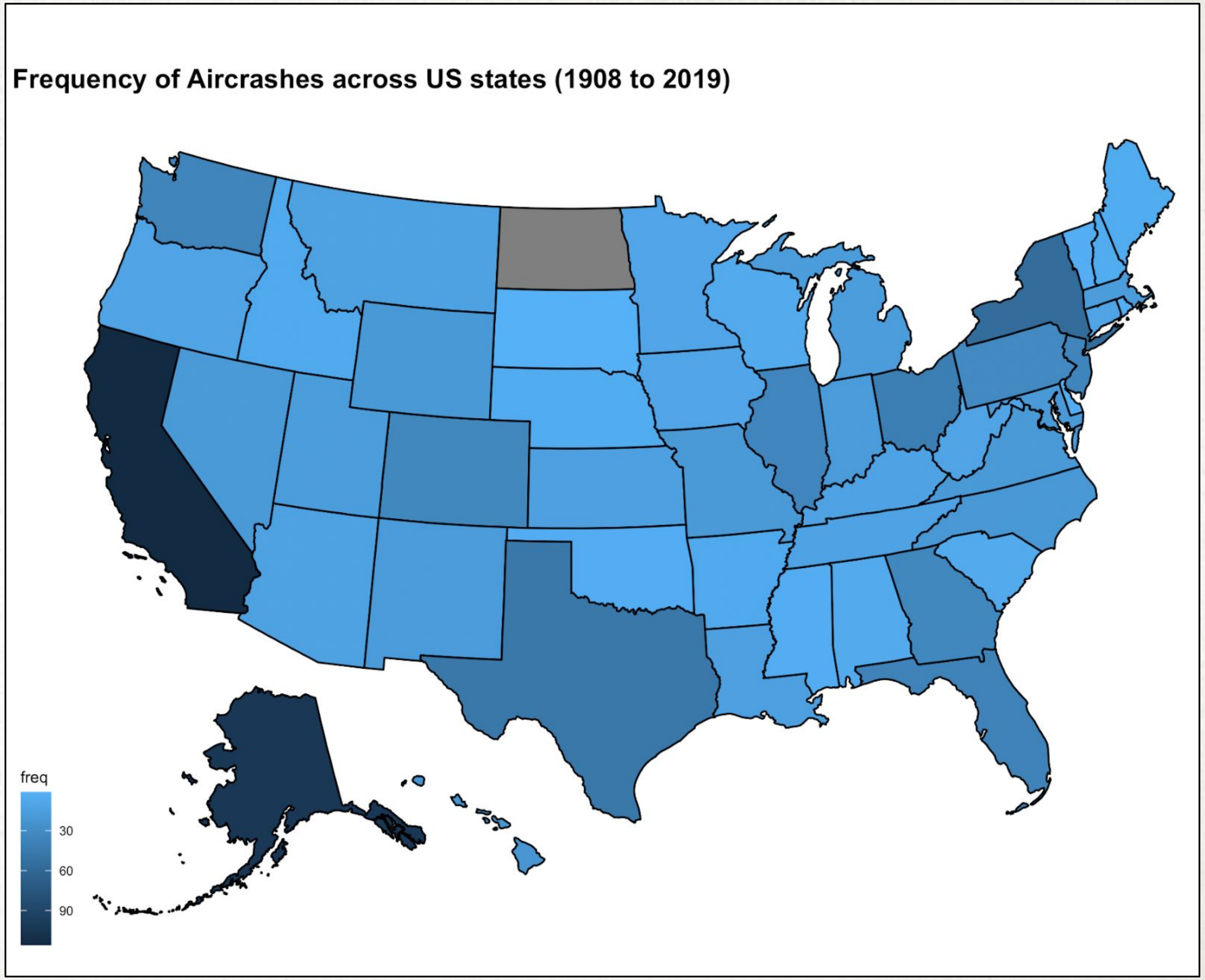


ANALYSES - Geographic Crash Analysis

From the data set the following are the frequency of air crashes across the various US States. Clearly California has the highest frequency of crashes, followed by alaska. There are various factors attributed to this cause like unstabilized weather patterns / Occurance of world war in those regions etc.



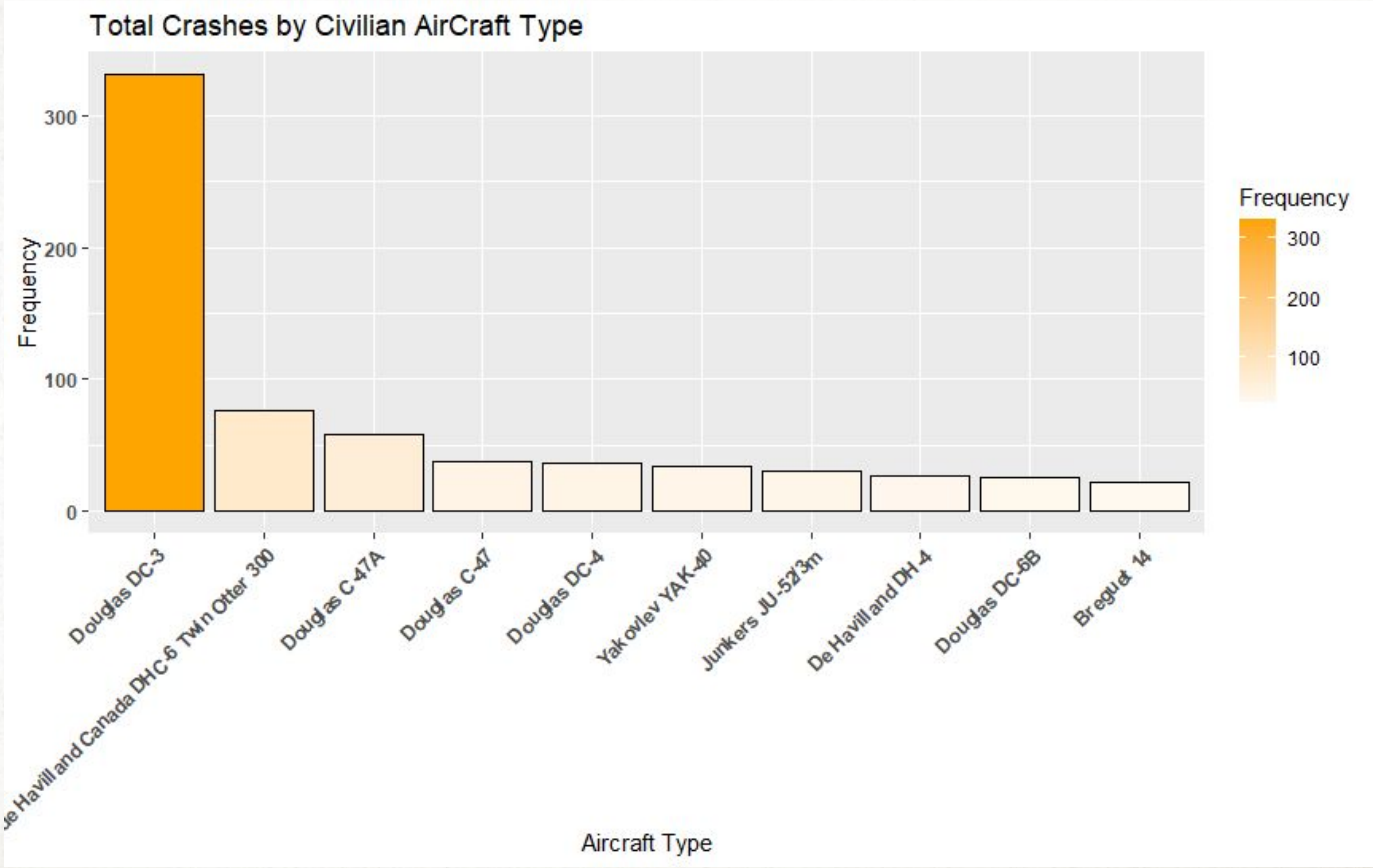
	x	freq
1	california	116
2	alaska	103
3	new york	56
4	texas	47
5	ohio	41
6	florida	38



Frequency of Aircrashes across US states (1908 to 2019)

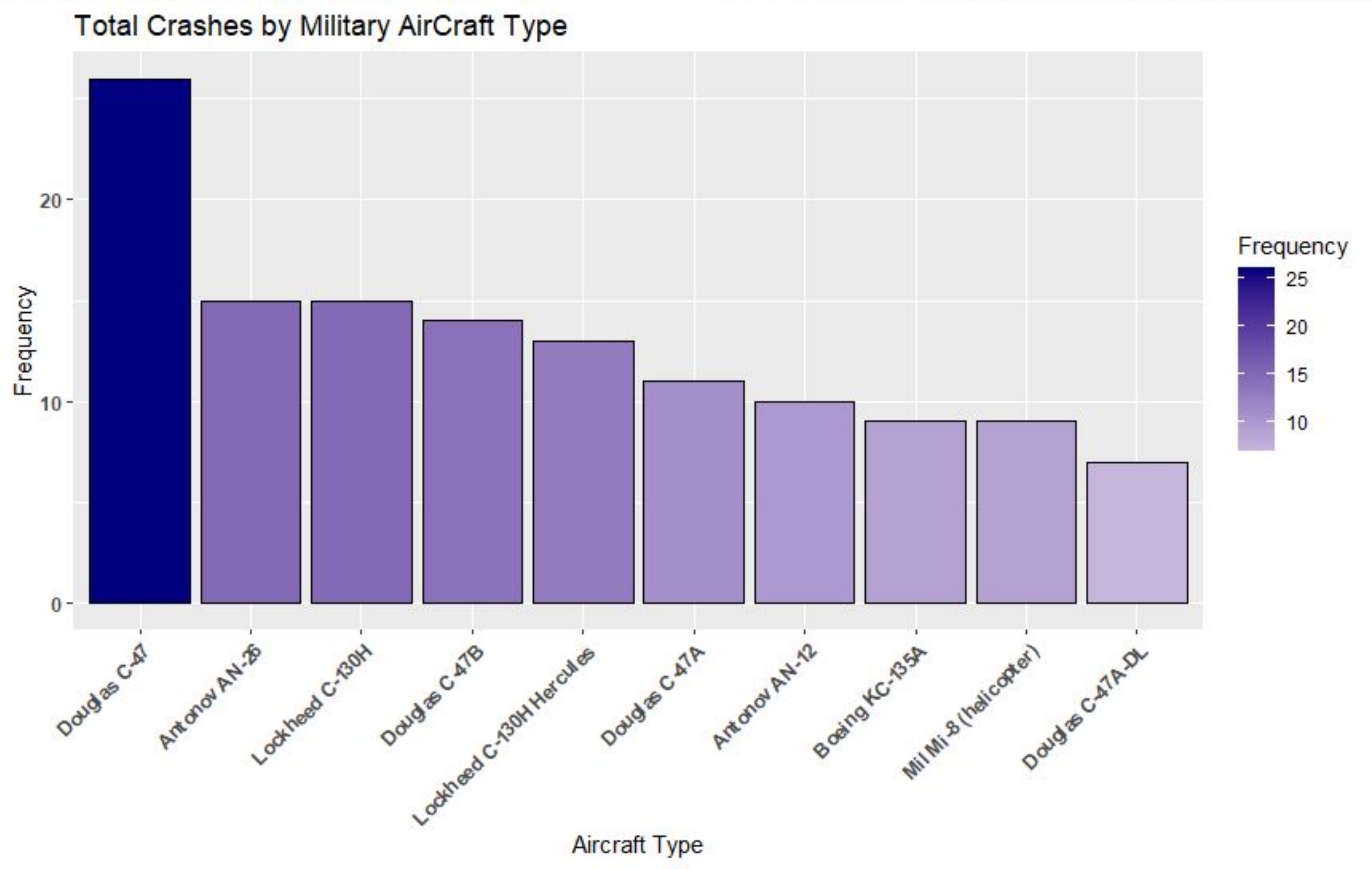
ANALYSES - Crashes by Aircraft Type

This bar chart shows the top ten Civilian Aircraft Types with the highest frequency of crashes as defined in the dataset over the course of the time. Clearly Douglas-D3 aircraft types founded in 1930 has by far the highest frequency of crashes The cause of Douglas-D3’s high crash rate could potentially be explained by its inadequate size and slow speed.



ANALYSES - Crashes by Aircraft Type

This bar chart shows the top ten Military Aircraft Types with the highest frequency of crashes as defined in the dataset over the course of the time. Clearly the Douglas C-47 aircraft types founded in 1941 has by far the highest frequency of crashes followed by Antonov AN-26 aircraft types used by the Serbian Air force. The cause of Douglas C-47’s high crash rate could potentially be explained by its limited maneuverability.



ANALYSES - DATA MODELING & VISUALIZATIONS INTRODUCTION/PREPARATION

Introduction:

K-means is a type of unsupervised learning that partitions observations into clusters based on certain similarities. You'll define a target number *k*, which refers to the number of centroids you need in the dataset. A centroid is the imaginary or real location representing the center of the cluster. Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares. The summary section of the dataset included reasons for the crashes with some being a paragraph long.

Data preparation:

Most of the preparation was done in the early stages of data cleansing. However, tokenization was necessary to partition the crash summary section apart into two terms. Two terms provided more meaningful results.

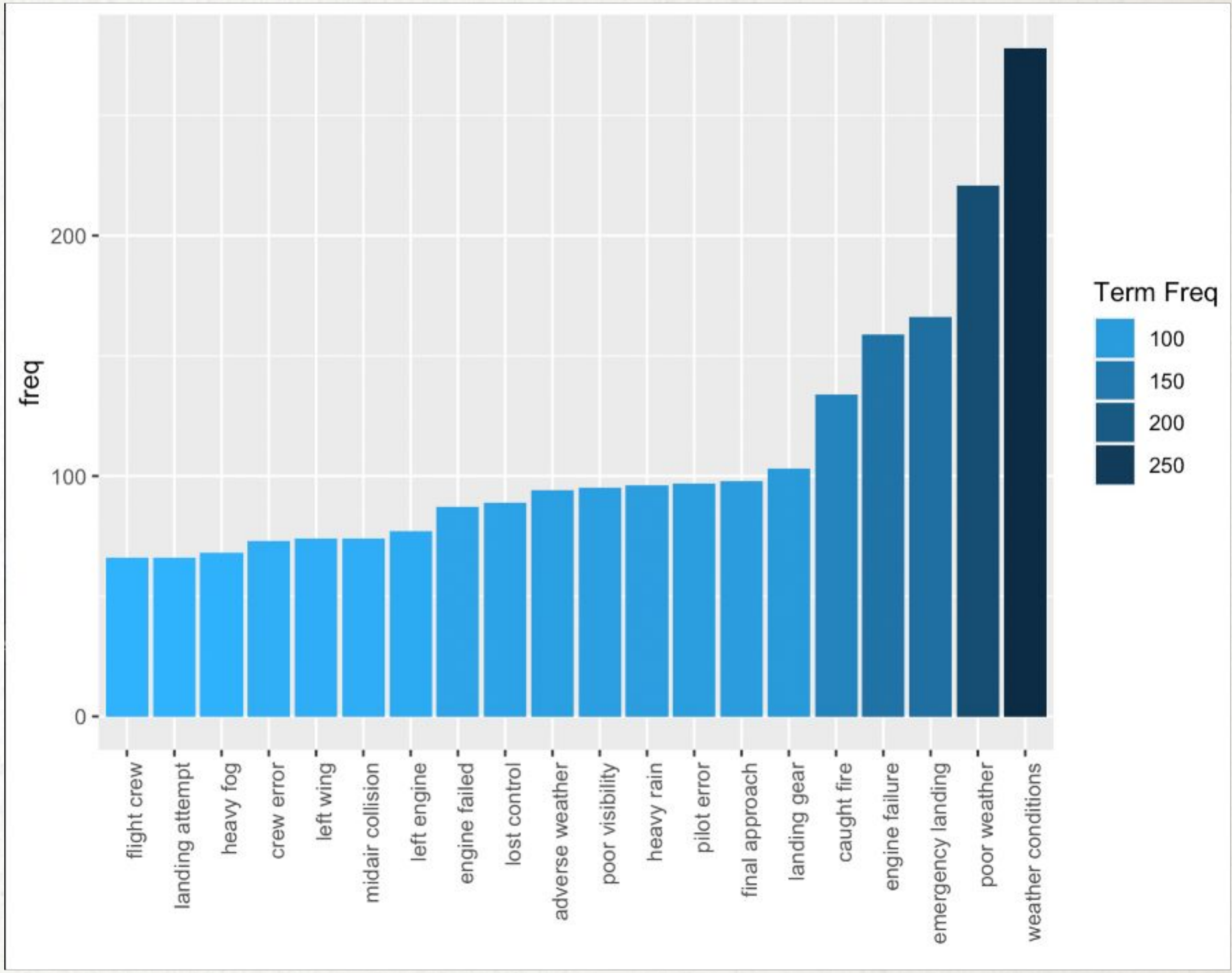
Model interpretation:

The Gap Statistic method was used to determine the optimum number of clusters to use. The basic idea of the Gap Statistics method is to choose the number of K (clusters), where the biggest jump in within-cluster distance occurred, based on the overall behavior of drawn samples. This method was computationally expensive but provided great results. The Dim1 and Dim2 is a form of principal component analysis, where the fatality variables (crew, passengers, ground) were "projected" into Dim1 and Dim2. Adding these two together we get 95.5%, which means together they explain 95.5% of the variation in the dataset.

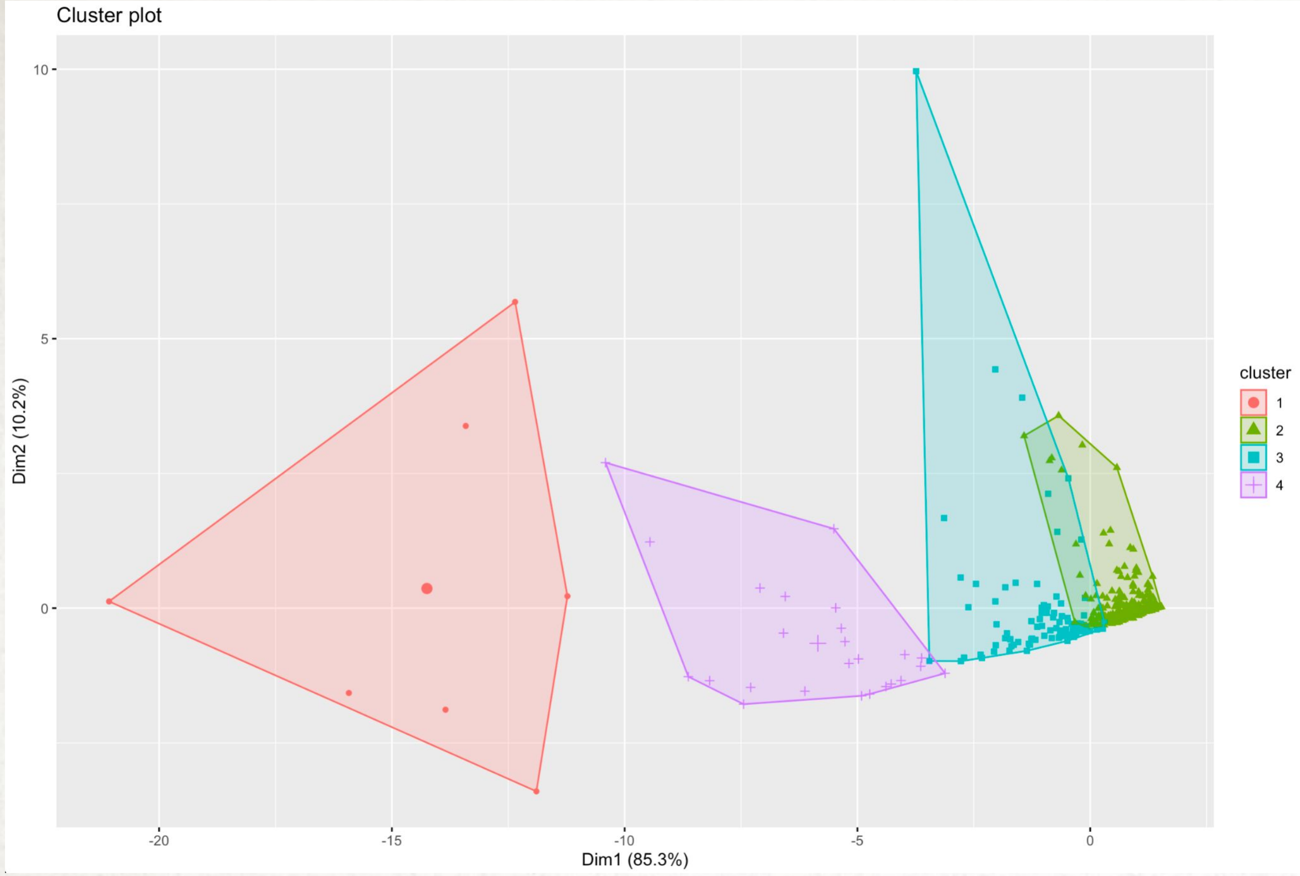
ANALYSES - Crash Reasons (Text Mining)

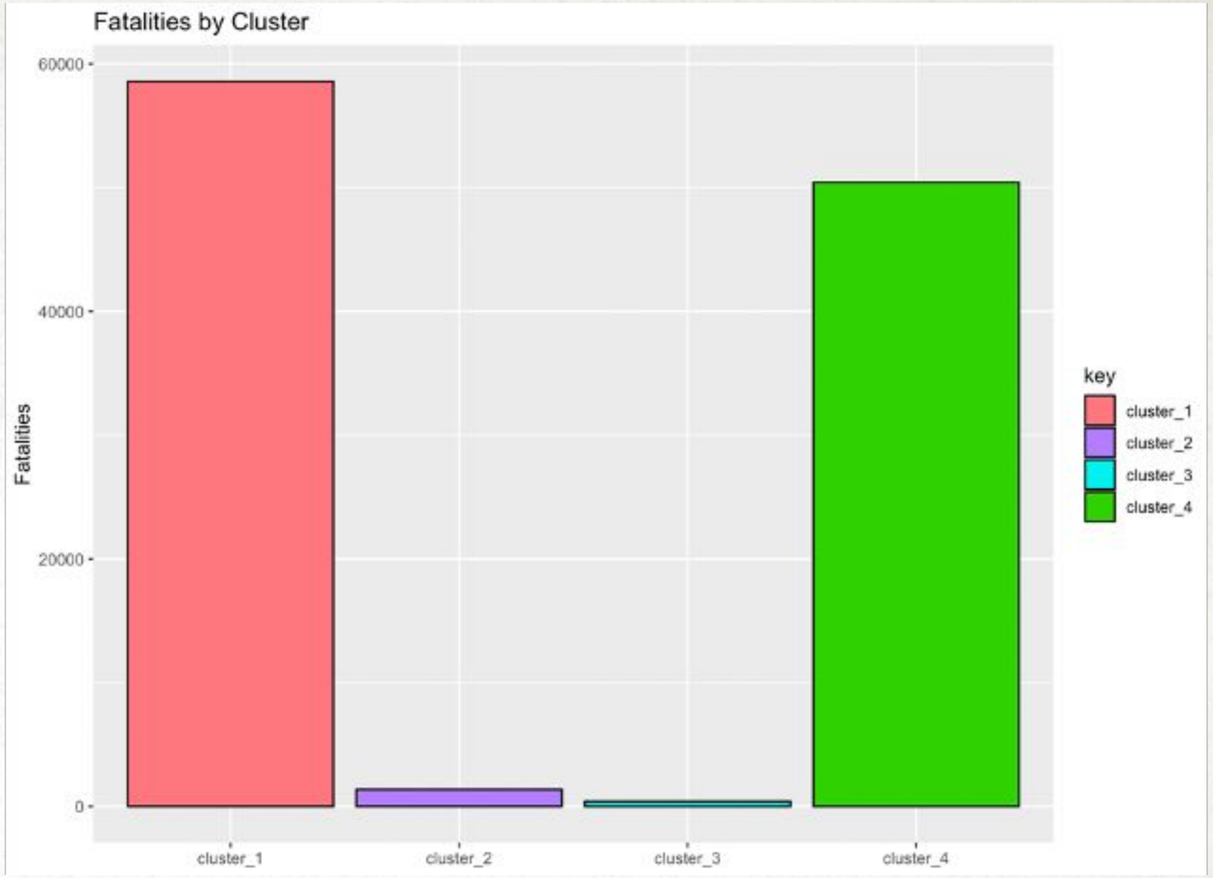
This bar chart presents the frequency of two word occurrences, which is called a bigram (N-gram analysis). Bigrams were utilized due to the lack of meaningful results when analyzing just one word or more than two. Cleary, weather was a significant reason for crashes between 1908 and 2019. Out of the top 20 reasons listed, seven were weather related followed by engine related causes.

It should be noted that there is some overlap with the terms. For example, some emergency landings were due to weather related events, engine failure, and pilot error.



ANALYSES - Crash Reasons (Text Mining continued…)



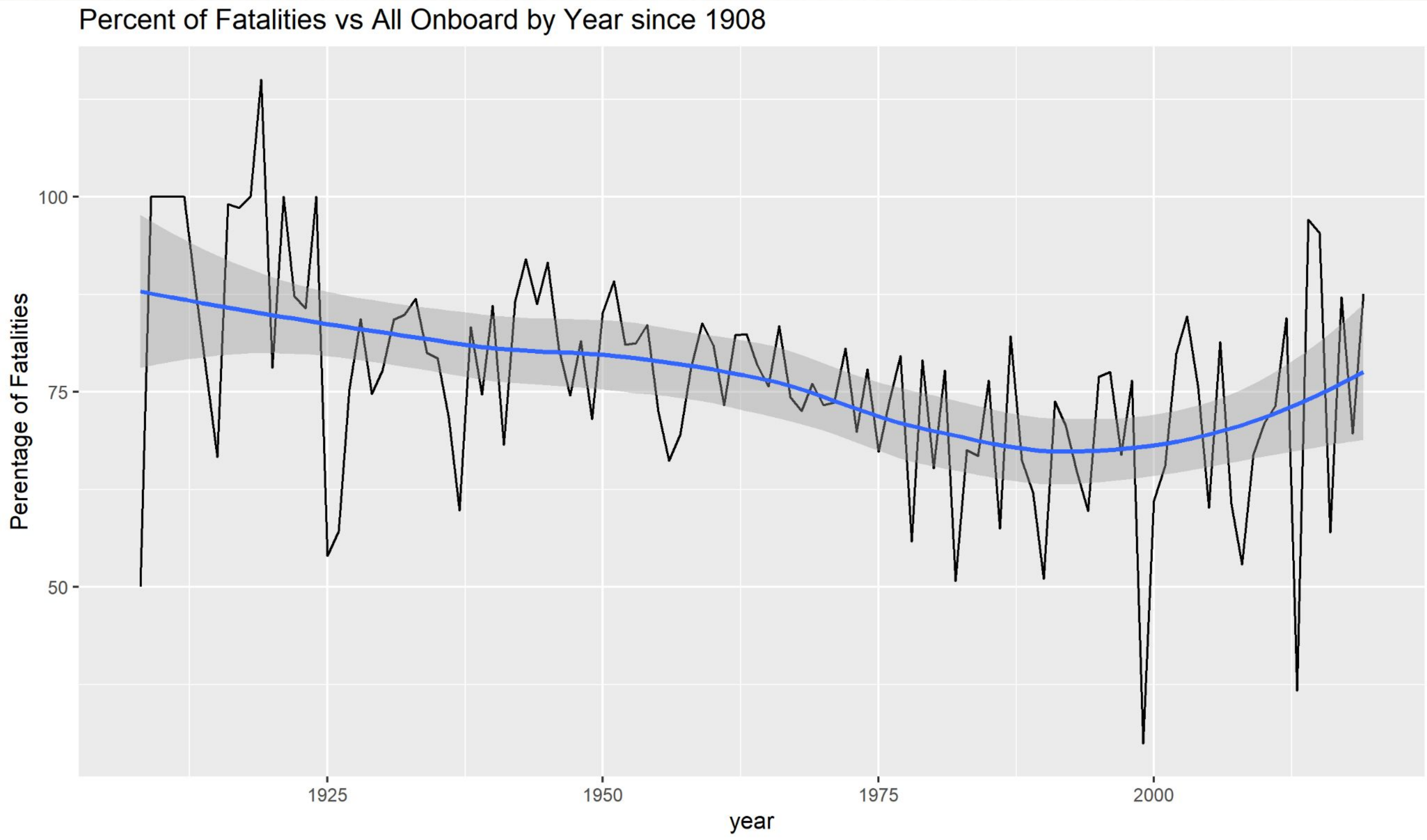


The graph to the left is the result of a K-means process, which is a type of unsupervised machine learning that was used for text mining. This process forms clusters, which is a collection of data points aggregated together based on similarities. The summary section of this dataset included reasons for the crashes with some being a paragraph long. Four clusters were formed based on similar text patterns with cluster 1 & 4 having the most fatalities.

- Cluster 1: Total Fatalities = **58,579**
  TopTerms=Weather, Emergency Landing, Engine Failure
- Cluster 2: Total Fatalities = **1,370**
  Top Terms= ATC Error, Atlantic Ocean, Residential Neighborhood
- Cluster 3: Total Fatalities = **407**
  Top Terms= Lost Altitude, Structural Failure, Air Missiles
- Cluster 4: Total Fatalities = **50,419**
  Top Terms=Pilot Error, Crew Error, Mid-air Collision

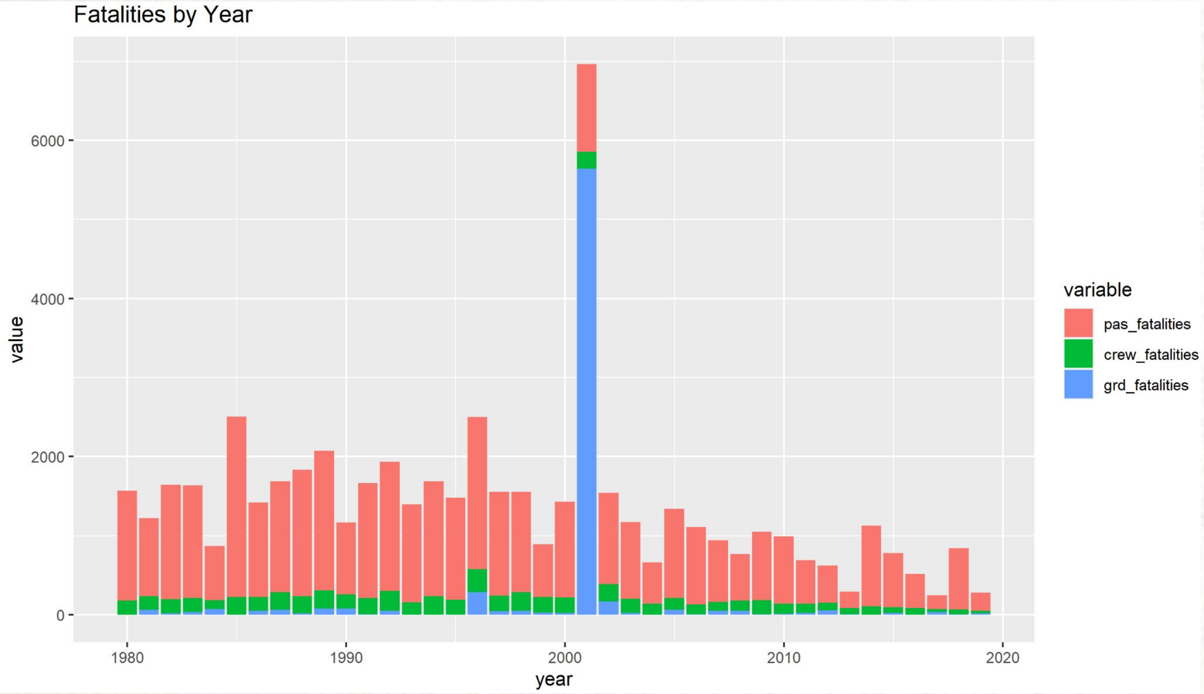
ANALYSES - Fatality Analysis

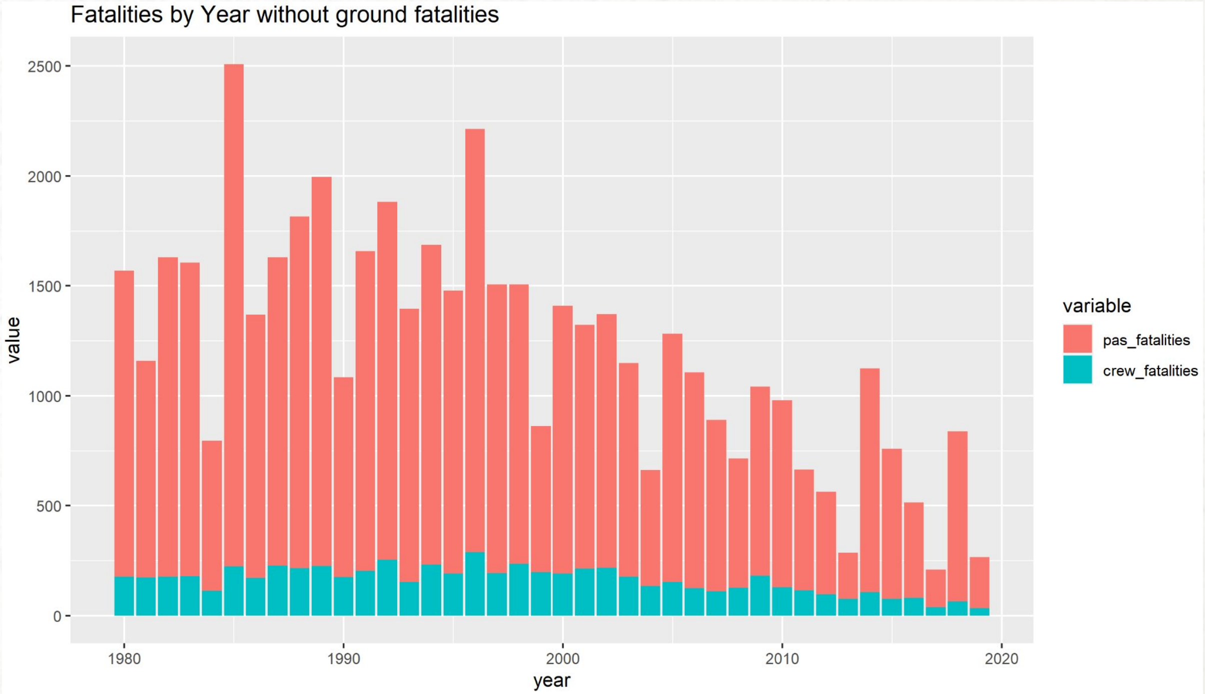
The percent of fatalities vs all onboard since 1908 ranges between ~65% and 88%. There had been a steady decline through the 1980s but noteworthy is that the percent began increasing from the 1990s until present.



ANALYSES - Fatality Analysis

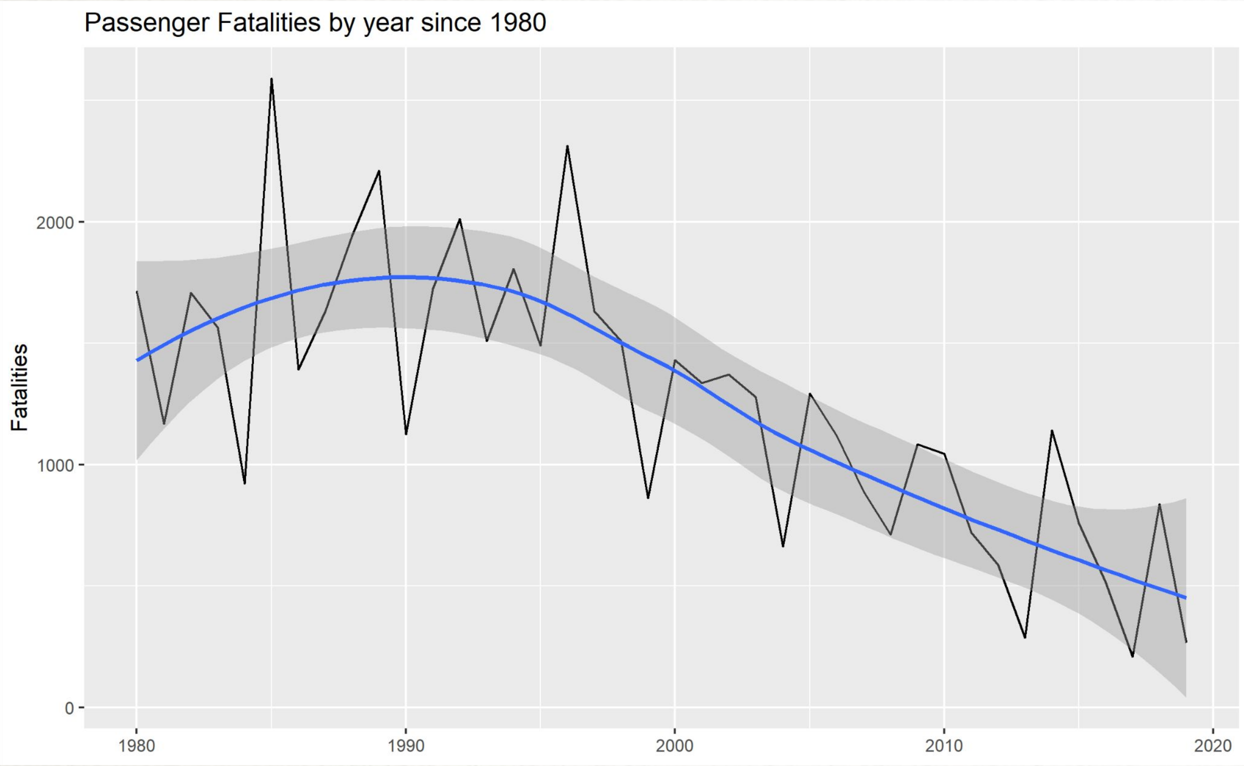
Fatalities are categorized in in terms of passengers, crew, and ground casualties. The vast majority of fatalities is typically passengers as would be expected. That said, ground fatalities were significantly higher in 2001 as the September 11th flights are included in our dataset (see the left stacked bar chart). Removing this anomalous event shows the breakdown without ground fatalities (see the right stacked bar chart).

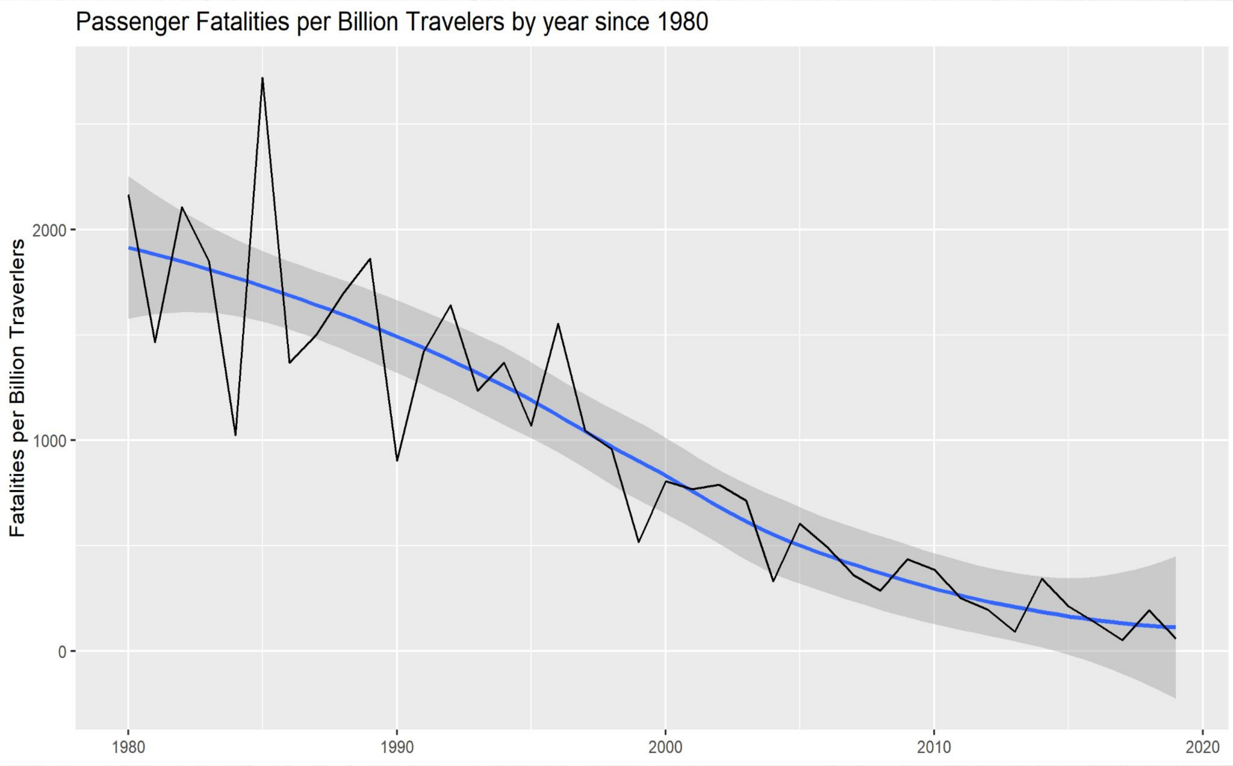




ANALYSES - Fatality Analysis

Fatalities increased from 1980 to 1990 and then started to decline (see the line chart on the left). However, based on increase in passenger volume, the number of fatalities per billion travelers has steadily declined (see the line chart on the right). Notably, 1985 was a particularly bad year as August of that year was "commercial aviation's deadliest month for passengers and crew" (see citation below).

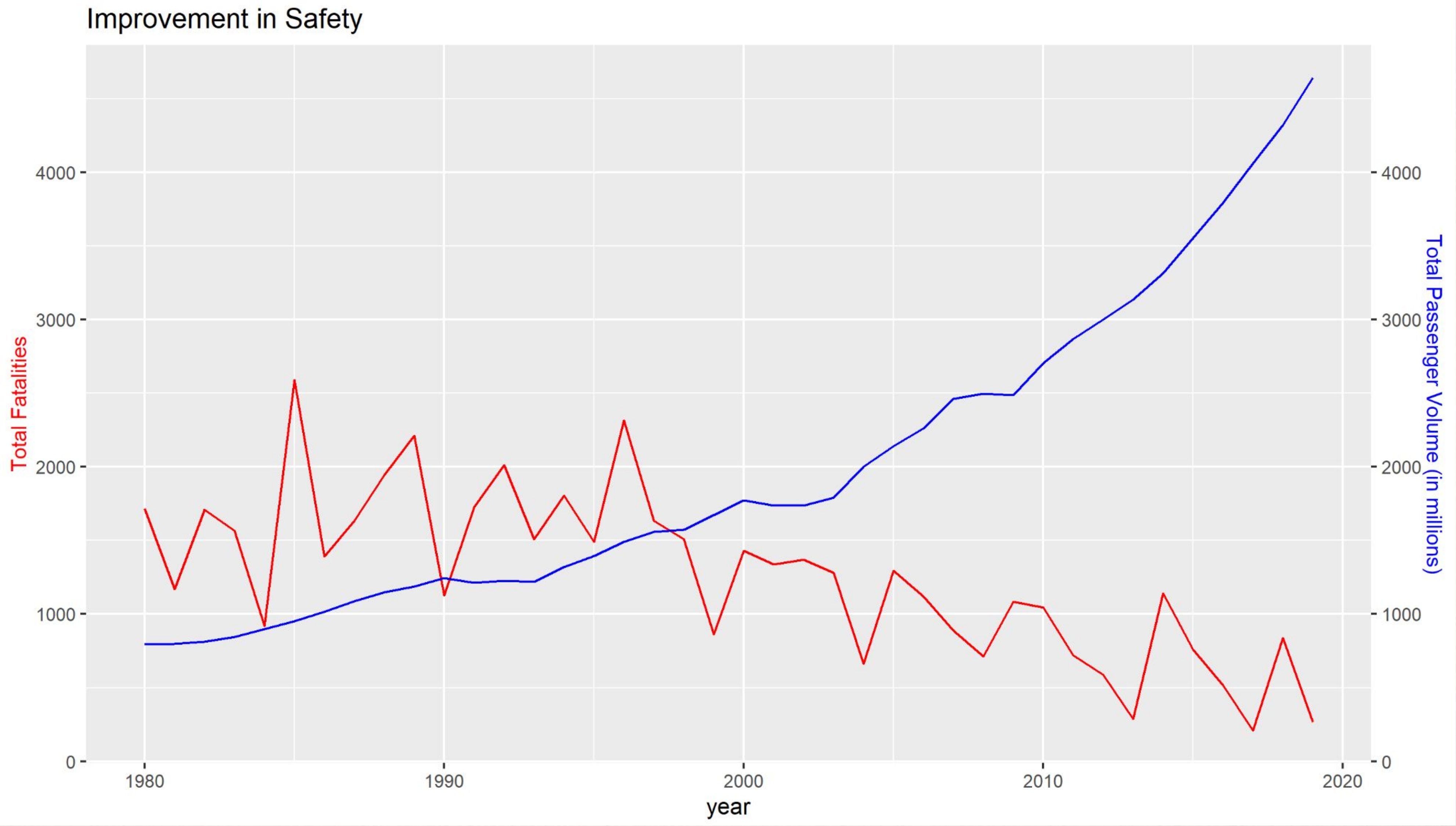




NOTE: From Wikipedia - August 1985 remains commercial aviation's deadliest month for passengers and crew (a distinction from the non-passenger fatalities of the September 11, 2001 attacks) in history.

ANALYSES - Fatality Analysis

Improvement in safety is ***OBVIOUS*** as fatalities are down while at the same time, passenger volume has increased exponentially since 1980.



PLANE CRASHES SINCE 1908

Table of Contents

Introduction

Data Cleansing/Munging

Analyses

Conclusions

Appendix
- R-code
- Raw Data Sets

Conclusions

- The main causes of crashes were weather related events, engine failure, and pilot (crew) errors.
- While crashes and fatalities are tragic events, based on an exponential growth in passenger volume since 1980, safety has dramatically increased.
- Civilian airlines crash more often than military air forces, and Aeroflot has significantly more crashes than other operators.
- Given its small population, Alaska has the 2nd most crashes just below California (the state with the highest population).

Bottom Line:

DON’T fly in Alaska during the winter on Aeroflot!

PLANE CRASHES SINCE 1908

Table of Contents

Introduction

Data Cleansing/Munging

Analyses

Conclusions

Appendix

- R-code
- Raw Data Sets

APPENDIX 1 – R-CODE

- See attached file

APPENDIX 2 - RAW DATA

- See attached file